CU12-H03125-62-P01445

THE AMERICAN ECONOMIC REVIEW

VOLUME LX

THE AMERICAN ECONOMIC ASSOCIATION

Executive Office: Nashville, Tennessee
Editorial Office: Brown University, Providence, Rhode Island

CONTENTS OF ARTICLES AND COMMUNICATIONS

CONTENTS OF PAPERS AND PROCEEDINGS

CONTRIBUTORS TO ARTICLES AND COMMUNICATIONS

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

THE AMERICAN ECONOMIC REVIEW

June 1970

VOLUME LX, NUMBER 3

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

Articles

Communications

THE AMERICAN ECONOMIC REVIEW

September 1970

VOLUME LX, NUMBER 4

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

Articles

Communications

Editor's Announcement:

The Sixty-Seventh List of Doctoral Dissertations in Political Economy will Appear in the December, 1970, issue of the REVIEW.

THE AMERICAN ECONOMIC REVIEW

December 1970

VOLUME LX, NUMBER 5

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

Articles

Communications

ALEXANDER GERSCHENKRON
DISTINGUISHED FELLOW
1969

Nullum ferre scribendi genus non tetigit,
Nullum quod tetigit non ornavit.*

This epitaph of Dr. Johnson for Oliver Goldsmith applies most suitably (with the usual allowance for hyperbole) to Alexander Gerschenkron. He is a profound and subtle scholar of the economies, the societies and the languages of Europe—East and West. He has illuminated in writings and in index numbers the complex nature of industrial growth and agrarian change in modern history. Most importantly, he has given to twenty-odd graduate classes and workshops of economists at Harvard a taste for the use of economics and history in combination, or at least in close proximity in their work, and a distaste for their misuse. A powerful lecturer, a hard task-master, a severe critic, and a warm and original mind and personality, he has infused into our profession by precept and example the finest traditions of continental scientific and humanistic scholarship and has provided us with some notable examples of both.

* There is hardly any kind of writing that he has not touched and nothing which he has touched that he has not adorned.

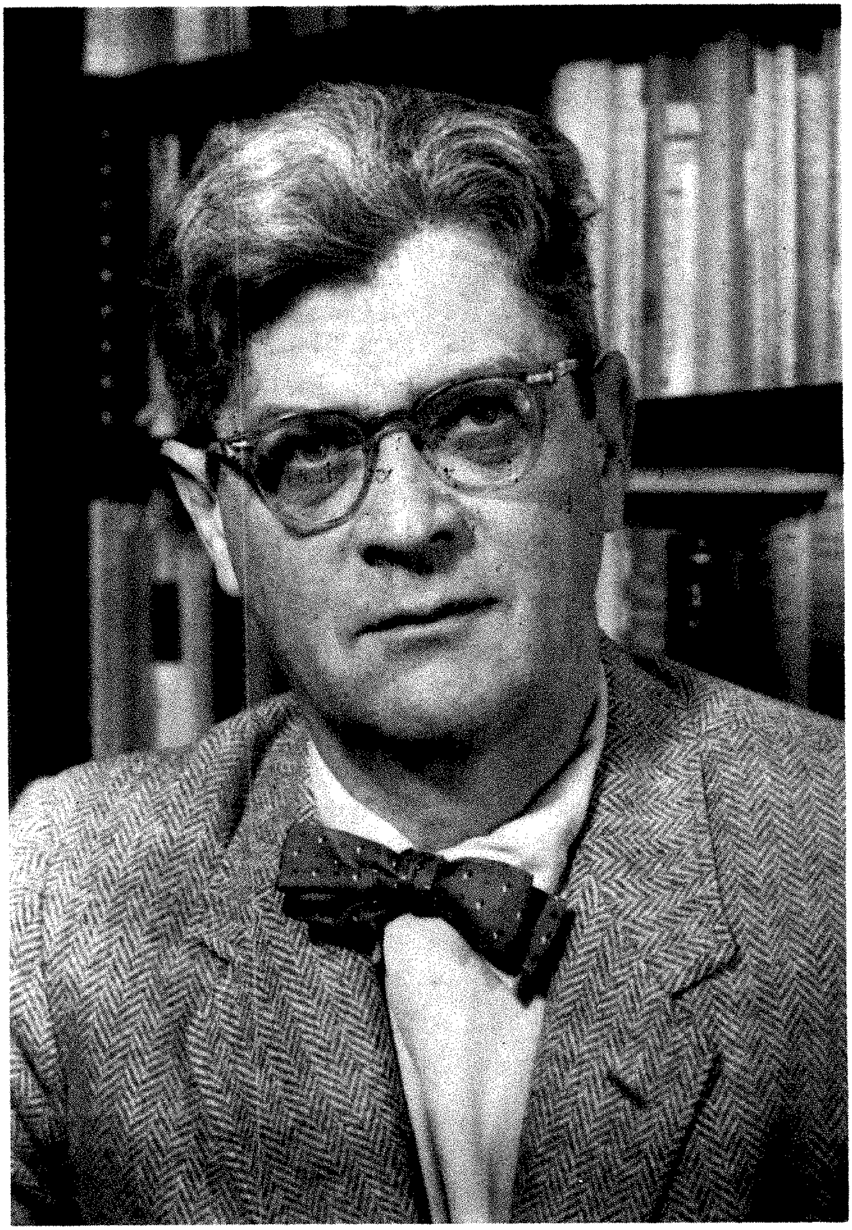
DISTINGUISHED FELLOWS
(In addition to Past Presidents)

1966
Abba P. Lerner

1967
Jacob Marschak

1968
Lloyd A. Metzler

1969
Ludwig von Mises



Alexander Gerschenkron

Optimal Departures From Marginal Cost Pricing

By William J. Baumol and David F. Bradford*

The need for this paper is a paradox in itself and indeed it might be subtitled: *The Purloined Proposition or The Mystery of the Mislaid Maxim.* For the results which it describes have appeared many times in the literature and have been reported by most eminent economists in very prominent journals. Yet these results may well come as a surprise to many readers who will consider them to be at variance with ideas which they have long accepted.

The proposition in question asserts that, *generally, prices which deviate in a systematic manner from marginal costs will be required for an optimal allocation of resources, even in the absence of externalities.* The reason for the difficulty into which marginal cost pricing is likely to fall is rather well known. What is not widely recognized is that there exists a highly sophisticated and well-developed body of literature indicating what should be done in such circumstances.

To see how the problem arises, consider an economy in which all industry has been nationalized and in which the central planning agency is dedicated to the maximization of social welfare. Suppose, accordingly, that it is decided to set all prices *equal* to marginal costs and to make up any deficits by subsidy out of the governmental treasury. If these funds are derived by excise taxes this is obviously a decision to make some prices depart from marginal costs after all. Or if it is obtained by an income tax, it is the price of labor which is forced away from its marginal cost. Any tax, except a Pigouvian poll tax—which might perhaps more felicitously be called an "inescapable tax"—will unavoidably affect some price. There is no way out of it. Any level of tax revenue which the government is determined to collect, whether as a means to make up a deficit resulting from a marginal cost pricing arrangement or for any other purpose, must in practice produce some price distortion.

Once this difficulty is recognized it becomes clear that one is dealing with a problem in the area of the second best. We are now faced with a problem involving maximization *in the presence of an added constraint.* Resource allocation is to be optimal under the constraint that governmental revenues suffice to make up for the deficits (surpluses)[1] of the individual firms that constitute the economy.

* The authors are members of the department of economics, Princeton University. We are grateful to the National Science Foundation whose grant greatly facilitated completion of this paper. We must also express deep appreciation to our colleagues, Charles Berry, Stanley Black, William Branson, and W. Arthur Lewis, and to Ralph Turvey, whose comments and suggestions helped enormously in the development of its ideas.

[1] Presumably, as a practical matter, there is an asymmetry between the problems posed by surpluses and those resulting from deficits. The real difficulty arises when we try to collect resources to cover the deficits without distorting consumers' and producers' choices. Since it is not possible in practice to levy lump sum taxes whose magnitude is independent of the decisions of those who pay them, we are forced to consider second-best solutions to the tax problem. A surplus can presumably be distributed more easily in a "lump sum" manner. One can, for example, distribute shares in future surpluses to all members of the economy on the basis of their incomes at some point in the past, or one can simply divide them equally among all individuals.

It should be emphasized also that the deficit problem need not arise only in the case of decreasing costs where

The theorems which have been developed to deal with this issue derive added interest from the fact that they are (virtually) the only concrete prescriptions for any second-best problem which have so far appeared in the literature. The discussion of this paper differs from the earlier writings in several ways. First, it deals with an important case not explicitly covered in most of the other papers in the area: quasi-optimal pricing when commodity prices throughout the economy are all adjusted as far as is possible. Second, it attempts a simplified exposition which is, of course, possible only at a cost in terms of loss of generality. Third, it brings together, explicitly, all three strands of the discussion: the welfare theoretic, the regulatory, and the public finance contributions. Finally, as far as we know, it offers the first overview of the extensive literature that has grown up in the area.

I. *Sources of the Analysis*

The preceding formulation of the problem suggests why the results described in this paper are not widely recognized. For they arise not out of the literature of welfare economics but from two more specialized fields—the theory of taxation and the analysis of public utility regulation.

The connection of the analysis with the theory of taxation is easy to understand. For the investigation, as it has just been described, is tantamount to a study of a system of optimal excise and income taxes. Only a small modification in our statement is needed for this purpose: instead of requiring the governmental revenue to equal the (algebraic) sum of the deficits of the firms constituting the economy, the revenue constraint can be generalized to require the tax system to bring in whatever revenues the government has decided it needs. The taxes which are optimal subject to this more general revenue requirement constraint then can be shown to follow the same necessary condition as that which replaces the marginal cost pricing rule in our previous discussion.

Similarly, there is a direct line between the more general welfare analysis and the theory of public utility regulation. For one reason or another, various public or regulated private monopolistic or partially monopolistic enterprises are operated under what amounts to a fixed profit constraint. Examples are turnpike authorities (required to cover costs and pay back invested capital), and regulated electricity, water, and telephone systems. The enterprises cited as examples are all firms that market a number of commodities. The turnpike sells easy driving between many different pairs of points, and, perhaps more important, at various times of day. Similarly, the telephone system sells ordinary telephone services for various uses over various distances and at various times, and, in addition, sells private wire, teletype, television signal transmission, and other services related to, but not identical with, telephone services as usually conceived. In each example the demands for the different services are very obviously strongly interrelated, and they are obviously produced under conditions involving common costs.

If we assume that the net revenue allowed these enterprises is less than the amount profit maximization would yield, some degree of freedom is introduced into the pricing of the commodities produced.[2] A wide variety of output combinations and

the revenue yielded by marginal cost pricing will fall short of the total costs of the firm. For example, in practice revenue requirements of the firm are often based on historical accounting cost figures which management feels it must recover. Similarly, governmental tax requirements may exceed any profits obtained from nationalized firms subject to diminishing returns when they price at marginal costs.

[2] Of course, some amount of freedom is also introduced into the choice of method of production. We assume throughout that good will, pride of service, patriotism or the shrewdness of regulators assure that whatever output combination is chosen is produced at minimum cost.

sets of prices can then be chosen, each of which would just meet the net revenue constraint. The problem, then, is to determine which of these is optimal (second best) from the point of view of the use of resources to serve consumer desires. A similar problem obviously arises in a nationalized industry which is required to earn enough to cover its costs.

II. *The Nature of the Theorem*

In this section we shall describe the rules indicating whether a particular price-output combination that satisfies the profit constraint is socially optimal, i.e., whether from the point of view of the economy it yields the most effective allocation of resources permitted by the constraint.[3] A solution to this constrained maximum problem will be called "quasi-optimal," because it is a second-best solution forced upon us by the revenue requirement.

A problem such as ours is usually framed in terms of the determination of appropriate output levels, but it can also be treated in terms of the choice of a price set for the outputs. Given the relevant demand functions, the choice of prices is tantamount to the choice of (salable) output levels. The two problems are, of course, formally identical, but, as we shall show, the pricing decision approach avoids a number of difficulties and yields a surprisingly simple optimality rule. In most conventional and familiar terms this rule asserts that Pareto optimal utilization of resources in the presence of an absolute profit constraint requires (considering substitution effects alone) that all outputs be reduced by the same proportion from the quantities that would be demanded at prices equal to the corresponding marginal costs. The rule takes an even simpler form in the event cross elasticities of demand are zero. It then requires that each price be set so that its percentage deviation from marginal cost is inversely proportionate to the item's price elasticity of demand. According to this result, the social welfare will be served most effectively not by setting prices equal or even proportionate to marginal costs, but by causing unequal deviations in which items with elastic demands are priced at levels close to their marginal costs. The prices of items whose demands are inelastic diverge from their marginal costs by relatively wider margins.

This result is surely not immediately acceptable through intuition. It strikes us as curious, if for no other reason, because it seems to say that ordinary price discrimination might well set relative prices at least roughly in the manner required for maximal social welfare in the presence of a profit constraint. Since the objective of the analysis can be described as the determination of the optimally discriminatory set of prices needed to obtain the required profit, some degree of resemblance is perhaps to be expected. The case studied here is, thus, in a sense the obverse of the problem of profit maximizing price discrimination, and while the two solutions bear some qualitative resemblance, it can be shown that they may in fact differ substantially in quantity.

The theorem can be reformulated and generalized in a number of ways. Instead of utilizing the producers' and consumers' surplus concepts with all their theoretical limitations, the analysis can be framed in terms of the Hicksian compensating variation. Or a Pareto optimality approach can be utilized, both of these procedures obviating any need for interpersonal utility

[3] It is not suggested that profit-constrained firms will by themselves tend to institute such optimal policies. When profits are eliminated or reduced in importance, other objectives presumably guide decision makers. For one thing, they may be able to afford stupidity in production decisions as pointed out by Harvey Leibenstein. A well-known alternative goal is sales maximization, the implications of which have been discussed *ad nauseam* by one of the present authors. Other possibilities have been examined by Herbert Simon, Robin Marris, Oliver Williamson, Armen Alchian, and Reuben Kessel and by others. For a bibliography and an excellent discussion see Williamson.

comparisons. Similarly, the theorem has been extended to cover cases involving nonzero cross-elasticities of demand, to deal with input prices and the prices of intermediate goods, etc. Of course, once this is done the result loses some of its simplicity and the preceding statement requires considerable modification. But even then, surprisingly simple versions remain possible, as we shall see.

III. *The Formal Theorem*

In previous discussions, the basic theorem has been stated in a number of different ways. The objective itself has been described alternatively as: a) maximization of the sum of consumers' and producers' surpluses; b) determination of a set of prices from which it is *not* possible to change in a way that permits the gainers to compensate the losers; and c) maximization of the level of satisfaction of any one individual, given the utility level of each other individual (Pareto quasi-optimality). Each of these maximizations is, of course, constrained by the revenue requirement.

We will discuss the following four variants of the theorem, each of which gives a set of necessary conditions for quasi-optimal pricing:

1) If prices are quasi-optimal the ratio between the marginal profit yields of unit changes in the *prices* of any two goods will be equal to the ratio between their output levels.

2) For each product, the deviation of the quasi-optimal price from marginal cost must be proportionate to the difference between the product's marginal cost and marginal revenue (i.e., its marginal social welfare cost must be proportionate to its marginal contribution to the profit requirement). This result holds only in the case where cross-elasticities are all zero.

3) For each product, the percentage deviation of quasi-optimal price from marginal cost must be inversely proportionate to its price elasticity of demand. This result also holds only where cross-elasticities of demand are zero.

4) Quasi-optimal prices must yield outputs that deviate by (approximately) the same proportion from those which would result from pricing at the marginal costs corresponding to the quasi-optimal output levels. This last form of the proposition, which is the variant most frequently encountered in the literature, is more general than the second and third versions of the theorem.

In demonstrating these propositions we shall utilize a comparatively straightforward manner of proof which has not previously appeared in the literature.[4] To facilitate the exposition we start with a partial equilibrium approach. That is, we demonstrate that socially optimal pricing by a multi-product monopolist operating under a profit constraint is described by the preceding propositions. The monopolist is assumed able to set the prices of his final good outputs, but to purchase inputs at prices which remain fixed throughout the analysis. It will be clear that his optimal policy is the same as that which would be induced by the imposition of excise taxes equal to the derived divergence between output price and the marginal costs which would obtain if the outputs were to be produced by perfect competitors. This establishes the correspondence between the monopoly regulation and the excise taxation interpretations of the theorems.

If the monopolist is taken to be very large (e.g., if substantially all of the economy is operated by the government), or correspondingly, if the commodities taxed account for a large share of total output, the assumption of fixed input

[4] To simplify the exposition we have dispensed with some theoretical niceties in the proof presented here, especially in the definition of the objective function. These are discussed and the objective function is examined in greater detail in our forthcoming paper.

prices must clearly be relaxed. This has been recognized and dealt with in the literature (for a simple statement see A. C. Pigou (1947, pp. 105–09); a sophisticated treatment is found in M. Boiteux 1956). We do not go into detail on this problem, but we do show in a subsequent section that the same line of reasoning used in our proof can readily be extended to a general equlibrium model.

The demonstration proceeds as follows:

Let $p_1, \ldots, p_n$ and $x_1, \ldots, x_n$ be the prices and outputs of the n commodities produced by the monopolist and, as indicated above, take as fixed the prices of his inputs. Let $Z(p_1, \ldots, p_n)$ be our (unspecified) measure of consumer benefit which is to be maximized subject to the profit constraint[5] $\Pi(p_1, \ldots, p_n) = M$ where the profit function has now been expressed in terms of prices rather than outputs. As usual in such constrained maximization problems, Z will be maximal subject to this constraint only if

$$(1) \quad \frac{\partial Z}{\partial p_i} = \lambda \frac{\partial \Pi}{\partial p_i}, \quad (i = 1, 2, \ldots, n)$$

That is the intuitively obvious first-order requirement that the marginal welfare gain from a given price change (price reduction) must be proportionate to its marginal profit cost.

[5] If the revenues permitted by regulation are based on total investment rather than being fixed in absolute amount, this constraint may require some modification. If (as is at least ostensibly the goal of regulation) gross profits are required to be no greater than the cost of capital, then net profits (over the cost of capital) will be zero and the constraint will hold as stated with $M=0$. If, however, the profit rate is permitted to exceed the cost of capital and the firm's total capital varies with output levels we may end up with M, the total profit permitted by regulation, itself varying with output levels. Our formal results remain unchanged if in the profit constraint $\Pi = R - C = M$ (where R = total revenue and C = total cost) we use C^* to represent $M+C$ and rewrite the constraint as $\Pi = R - C^* = 0$. We can then express all of our results in terms of C^* instead of C. Economic interpretations are, however, clouded by this modification.

We can make the preceding equation much more explicit by examining more specifically the nature of the consumer benefit function $Z(p_1, \ldots, p_n)$. While we will not be able to say much about that function itself, we can be rather specific about its partial derivatives, which is all we need for our present purposes.

We need a basis on which to estimate the gain to consumers from price-quantity data. A one dollar reduction in the price of a commodity will enable a consumer who is presently consuming x units of that commodity to continue buying exactly his original bundle of goods with a budget that is smaller by x dollars than his original budget. Thus x dollars is a lower limit to the relevant compensating variation since this is the least he would be willing to pay for that price change, and he would probably pay more since the x dollar reduction in his budget accompanied by the one dollar reduction in the specified price would certainly permit him to buy his original bundle and might perhaps enable him to buy one which he prefers but which he could not previously afford. Similarly, if an individual is currently purchasing the respective amounts $x_1, \ldots, x_n$ of our n commodities at prices $p_1, \ldots, p_n$, the amount he would pay to confront instead prices $p_1 + \Delta p_1, p_2, \ldots, p_n$ will be at least $-x_1 \Delta p_1$. John Hicks (p. 330 ff.) shows rigorously in this same general manner that the *rate* of consumer gain per dollar *increase* in the price of any commodity is precisely the negative of the amount of the commodity initially consumed, times one dollar. We may then take as our expression for the derivative of the benefit function the compensating variation corresponding to the price change

$$\partial Z / \partial p_i = -x_i$$

Hence, substitution into the previous equation (1) yields

$$(2) \qquad -x_i = \lambda \partial \Pi / \partial p_i$$

or, dividing by the corresponding condition for any other item j, we obtain

$$\text{(2a)} \qquad \frac{1}{x_i}\frac{\partial \Pi}{\partial p_i} = \frac{1}{x_j}\frac{\partial \Pi}{\partial p_j}$$

as the required condition for quasi-optimality.

Equation (2a), then, is the first of the four variants of the theorem described at the beginning of this section. This necessary condition for quasi-optimality requires marginal profit yields of price changes to be proportionate to output levels. It may perhaps claim the virtue that it is relatively operational. If the firm has some notion of the likely marginal profit yield of a change in any one of its product prices, management can readily check whether (2a) is approximately satisfied.

We can now proceed directly from (2) to a derivation of the other variants of the basic theorem. We use MR_i, MC_i, and E_i to represent, respectively, the marginal revenue, marginal cost, and price elasticity of demand of output i, and for simplicity assume that all cross-elasticities of demand are zero, though this is not necessary for the more general analysis, and is certainly not required for the preceding variant of the theorem.

Since[6]

$$\text{(3)} \qquad MR_i = p_i + x_i \partial p_i/\partial x_i$$

then

$$\partial\Pi/\partial p_i = (MR_i - MC_i)dx_i/dp_i$$
$$= (p_i + x_i\partial p_i/\partial x_i - MC_i)dx_i/dp_i$$

Substituting this into our basic condition (2) we have

$$-x_i dp_i/dx_i = \lambda(p_i + x_i dp_i/dx_i - MC_i)$$

or adding $(p_i + x_i dp_i/dx_i - MC_i)$ to both sides we obtain

$$p_i - MC_i = (1+\lambda)(p_i + x_i dp_i/dx_i - MC_i)$$
$$= (1+\lambda)(MR_i - MC_i)$$

This is the second form of the quasi-optimality theorem: the difference between price and marginal cost[7] should be proportionate to the difference between marginal revenue and marginal cost.

Next we may rewrite the preceding equation as

$$\text{(4)} \quad -\lambda(p_i - MC_i) = (1+\lambda)x_i dp_i/dx_i$$

or

$$\frac{p_i - MC_i}{p_i} = \frac{1+\lambda}{\lambda}\frac{1}{E_i}$$

since by the usual formula, $E_i = -(p_i/x_i)(dx_i/dp_i)$ which is the third form of the result.[8] Note that it implies that if all elasticities in question are equal, prices

[6] It is here that the zero cross-elasticities enter our exposition. In the more general case, of course, $MR_i = p_i + \sum_j x_j \partial p_j/\partial x_i$. In our forthcoming paper we have dealt in some detail with this more general case which clearly complicates the analysis but does not change its nature in any fundamental way.

[7] Some care must be exercised in interpreting the various forms of the theorem described here because marginal costs are not generally constant. Let $MC_i(q)$ represent marginal costs corresponding to the quasi-optimal output levels while $MC_i(M)$ represent the marginal costs corresponding to the equilibrium outputs when all prices are set equal to marginal costs. Then in our proposition we have $\Delta p_i = p_i - MC_i(q)$ and they are thus *not* necessarily equal to the $p_i - MC_i(M)$. That is, the propositions do not contrast quasi-optimal prices with the prices corresponding to a marginal cost pricing equilibrium as ordinarily conceived. They refer instead to the differences between the quasi-optimal prices and the marginal costs corresponding to the quasi-optimal output levels.

[8] It should be reemphasized that this form of the derivation deliberately ignores the supply side for the sake of expository simplicity. However, the analysis is easily extended to take supply into account, and this was done from the very beginning of the discussion, by Frank Ramsey and by A. C. Pigou in 1947. For example, Pigou states "Writing η (defined as negative) for the elasticity of demand in respect of pre-tax output, e_r for the corresponding elasticity of supply of the rth commodity and t_r for the ad valorem rate of tax on it, it can be proved that, in the conditions supposed (the optimum system of such taxes yielding a given revenue require) rates of tax . . . such that $t_r/(1/e_r - 1/\eta)$ has the same value for all values of r" (pp. 107–08).

should be set proportional to marginal costs.

For the final form of the proposition, write the deviation of price from marginal cost as Δp_i, i.e., set $p_i - MC_i = \Delta p_i$. Then equation (4) can be written after a minor rearrangement of terms as

$$\frac{dx_i}{dp_i}\Delta p_i = -\frac{1+\lambda}{\lambda}x_i$$

But the left-hand side of the preceding equation may be interpreted as an approximation to the change, Δx_i, in the ith commodity demanded which would result from a shift from the actual prices to current marginal cost levels. We thus obtain the last version of our theorem:

$$\Delta x_i = kx_i, \quad \text{where } k = (1+\lambda)/\lambda.$$

The more sophisticated discussions have generally emphasized this last form of the theorem, the assertion that quasi-optimal pricing requires (after compensation for income effects) a proportionate change in all purchases from the levels that would be observed if prices were set at marginal costs.[9] This interpretation has usually been preferred to the two variants that precede it because, as has been shown by M. Boiteux (1956), Paul Samuelson (1951), and others, this form of the theorem holds quite generally and does not depend on the simplifying assumptions needed to arrive at the two special forms of the theorem described before this last variant. Only our first relationship (2a), which requires the marginal profit yields of commodity price changes to be proportionate to their output levels, is of the same order of generality.[10]

IV. *Intuitive Rationale*

Though at first blush the rationale underlying the preceding propositions may not be entirely transparent, it is not too difficult to account for them intuitively. The last form of the theorem is perhaps the most helpful in offering us a grasp of the entire matter. We need merely think of the consequence of a deviation of price from marginal cost as a distortion of relative demand patterns. This immediately suggests the last form of the theorem for it, implies that the damage to welfare resulting from departures from marginal cost pricing will be minimized if the *relative* quantities of the various goods sold are kept unchanged from their marginal cost pricing proportions. If we accept this plausibility argument, then the elasticity form of the theorem clearly follows. If we propose to (say) contract the demands for A and B each by precisely k percent, then if A's demand is much more elastic than B's, clearly A's price must be raised by a substantially smaller percentage than that of B.

A bit more light is shed on the matter by a simple graphic discussion that follows rather loosely an analysis first presented by Ursula Hicks (p. 167 ff) and developed

[9] In accord with the comments of fn. 6, this third form of the theorem should be read as follows: a change in prices from their quasi-optimal levels to *the corresponding marginal cost levels* should reduce all outputs in (approximately) the same proportions.

If we estimate demand changes on the assumption that the slope of each demand curve is constant over the relevant range, our proposition is valid for large as well as for small price changes. It now asserts that the divergence between price and marginal cost must be such that the estimated percentage changes in demand resulting from a drop in all prices to the given marginal cost levels are the same for all commodities.

[10] Note that application of the theorem in any of its forms does not rule out a two-part or a multi-part tariff. On the contrary, it tells us how to determine the quasi-optimal value for each of the multiplicity of prices that composes such a tariff. For example, when the consumer pays one charge for the rental of a telephone and another for each of his calls, this can be interpreted either as a two-part tariff or as a pair of prices for two distinct but related services each of which has a quasi-optimal price. This discussion is, of course, based on the observation that one cannot in reality impose a Pigouvian poll tax as one part of the multi-part tariff. Indeed, if such a lump sum tax were possible there clearly would be no need for the theorems under discussion. Marginal cost prices plus lump sum taxes could then satisfy the revenue requirement and yield an optimal allocation.

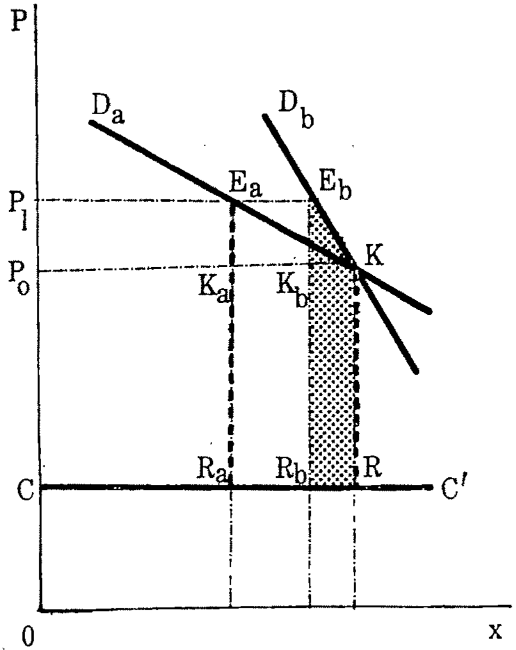


FIGURE 1

fully by William Vickrey in 1968. In Figure 1 consider the segments of two demand curves D_aK and D_bK through point K. Let us use the Marshallian measure of consumers' surplus to compare the psychic loss to consumers from a given price (tax) rise with the resulting increase in (tax) revenues. Then, if demand is in fact given by the less elastic curve D_bK and price rises from P_o to P_1, the loss of consumers' surplus will be $P_oKE_bP_1$. On the other hand, assuming for simplicity that marginal costs are constant at level OC, net revenue will be increased as a result of the price change by the quantity $P_oK_bE_bP_1$ minus R_bRKK_b. Thus the positive portion of the net revenue change, $P_oK_bE_bP_1$, may be considered to be offset by an approximately equivalent loss to consumers. This means that with the less elastic demand curve the rise in price will have caused a net reduction in net revenue plus consumers' surplus that is measured by the shaded area, R_bRKE_b.

Similarly, with the more elastic demand curve, D_aK, the rise in price from P_o to P_1 will decrease the sum of the revenue net of consumers' surplus by the greater area R_aRKE_a. In general, then (with marginal costs constant), a given price rise will exact a larger social cost in consumers' surplus not offset by increased revenues the more elastic the demand curve. That is essentially the reason for the theorem that calls for a relatively small deviation between the price and marginal cost of any commodity whose demand is comparatively elastic.

V. *A Simple General Equilibrium Model*

Our interest in conveying the central results of the literature discussed in our paper to a wide readership in an intuitively convincing fashion leads us to couch our argument of Section IV in terms of partial equilibrium analysis. It is desirable, however, to sketch out the way it all works in general equilibrium terms. A simple reinterpretation of the terms used in the derivation of the theorems will show that our proof is sufficient to establish the propositions in the special general equilibrium case in which there is only one input factor, labor, and all the productive activities are in the hands of the government. Accordingly, let $p_1, \ldots, p_n$, be the prices in terms of labor hours at which the n final outputs other than leisure, $x_1, \ldots, x_n$, all produced by the government, are sold to the consumer-laborers. The government is assumed to operate with a profit constraint in the form of a fixed number, M (which may be zero), of units of labor. In addition, of course, there is a production constraint, which we may write as[11]

[11] Notice that in putting this constraint in the form of an equality, we have ruled out production inefficiency. Although this is a plausible requirement and makes things easier, it is not always innocuous when more is meant by "consumer benefit" than attainment of Pareto optimality—the concept utilized here. See

(5) $$C = F(x_1, \ldots, x_n),$$

where C is the minimum labor input sufficient to produce the output vector, $x_1, \ldots, x_n$. The problem is to maximize "consumer benefit" subject to this production relation and the profit constraint,

(6) $$\pi = p_1x_1 + \ldots + p_nx_n - C = M$$

Now let us treat the government as an ordinary firm that imagines it can buy all the labor it wants at a price of one per unit. The additional labor it would have to buy to produce an additional unit of good is then the marginal cost of that good, both in the accounting sense and in the sense of its leisure opportunity cost, all other outputs held constant.

There will be, in general, a large set of price vectors compatible with both (a) clearing of the markets for goods and factors and (b) satisfaction of the government's profit constraint. It is useful to spell this statement out in somewhat more detail. If the aggregate labor sold is L, we can write the set of demand equations as

$$\begin{aligned} L &= L(p_1, \ldots, p_n) \\ x_1 &= x_1(p_1, \ldots, p_n) \\ &\cdots\cdots\cdots \\ x_n &= x_n(p_1, \ldots, p_n), \end{aligned}$$

where, necessarily (Walras' Law),

$$p_1x_1 + \cdots + p_nx_n = L$$

For a set of prices to be market-clearing, the government must buy exactly the specified amount of labor and produce exactly the specified quantities of goods. By (5) and (6) "feasible price vectors" are those which satisfy the profit constraint conditions

$$L - F(x_1, \ldots, x_n) = M,$$

or, equivalently,

$$p_1x_1 + \cdots + p_nx_n - F(x_1, \ldots, x_n) = M$$

The task is to find among these an optimal price vector, where we have taken this to mean a vector which consumers in the aggregate would be unwilling (after costless negotiation among themselves) to bribe the government to change. The solution to this problem is found in the propositions set forth in Section III above. The entire analysis in that section continues to hold when the profit constraint as just described is used in place of the usual partial equilibrium construct.

VI. *The Taxed and the Untaxed Sectors*

Now we are equipped with a genuine, if very simple, general equilibrium apparatus, but one which in most respects can be manipulated in partial equilibrium terms. The key to this is the form in which we have expressed the production relation (as a cost function), the fixing of the government's profit constraint in terms of labor, and the corresponding choice of labor as price numeraire (so that the price of the single input factor, the "all other prices" of partial equilibrium analysis, is fixed at unity). In this model the government can be taken to tax the output of producers of the n non-leisure goods (where the tax is the difference between price and marginal cost), and hence these goods make up the "taxed sector." The untaxed sector here may reasonably be described as the production of negative labor input to the taxed goods-production sector, i.e., the production of leisure. In a rather trivial sense, in the latter sector marginal cost and price are always equal at unity.

It is illuminating at this point to reconsider "our" results in light of the following comment by Abba Lerner, who derives two special rules for quasi-optimality, each applicable to a special case:

[One] rule [equalize tax elasticities] is

Peter Diamond and James Mirrlees for a proof that production efficiency will be called for under almost all "reasonable" assumptions about production, individual utility, and social welfare relations.

> appropriate where the shifting is only from a taxed to an untaxed sector, while the [other] rule [maintain proportionality of price to marginal cost] is appropriate where the shifting is only to another part of the taxed sector (as would be the case if all uses of resources could be, and were, taxed).
>
> What we need is a rule that is applicable whether the resources shifted by a tax remain in the taxed sector or go to the untaxed sector or are divided between the two sectors. [p. 285]

Documentation of the fact that a general rule for optimal excise taxation has long been available in the literature, and presentation of a simplified derivation of several forms of such a rule are precisely our objectives. It is easy to show that in those situations where Lerner recommends prices in proportion to marginal cost, our theorems do likewise; where Lerner recommends equalization of tax elasticities, so do our theorems; and where Lerner calls for something in between the two, our theorems provide an exact prescription.

In the case of the model which has just been described,[12] the statement that a tax change has as its only effect a shifting of resources from taxed to untaxed sectors is clearly equivalent to the assertion that the only output affected in the taxed sector is the one whose tax rate has changed. If the effect of varying any one tax is only to shift resources into the non-taxed sector then only own elasticities of the taxed goods, and not their cross-elasticities with one another, can be non-zero. In this case, Lerner and we (most directly in variant 3 of the theorem) conclude that the percentage deviation of price of any taxed commodity from marginal cost should be inversely proportional to its own price elasticity of demand.

We turn now to the next case where the alternative use of resources induced by tax changes can be said to be entirely in the taxed sector. This is obviously equivalent to the assertion that labor is perfectly inelastically supplied over a wide range of price vectors.[13] In this situation, Lerner and our theorems agree that prices proportional to marginal costs, i.e., relative prices equal to marginal rates of transformation between (non-leisure) commodities, is clearly optimal. This result follows easily for the case when cross-elasticities between (non-leisure) goods are all zero. Then, since a change in the price of a given good results in no change in either the purchases of other goods or of labor sold, it must lead to no change at all in labor expenditure on the good in question. That is, all goods must be demanded with equal, in fact, with unit elasticity. Our third form of the theorem states that in this case the relative deviation of price from marginal cost should be the same for all commodities.

More generally, admitting non-zero cross elasticities, the zero elasticity of labor supply means that any change in profit resulting from a price change comes purely from the change in labor cost of goods sold, since there can be no change in the

[12] Two things should be noted. First, when Lerner speaks of "tax elasticity" he is referring to what might better be called "own tax elasticity," that is, the relative change in the tax revenue (here, profit) collected on commodity x_i when the tax on x_i (divergence between price and marginal cost of x_i) is raised by one unit. The change in tax revenue collected on other commodities, which arises via cross-elasticity of demand, is not included.

Second, it should be observed that although the term "resources" is often not definable in general equilibrium models of the type described above (since choice is made from among alternative vectors of social aggregates of final goods, without reference to the underlying mechanics of production), in this simple case we may presumably equate them with "labor."

[13] Note that unless this inelasticity of supply results from some kind of artificial constraint, such as a legal maximum working day, or from a kink in the individuals' utility functions, commodity taxes will in this case *not* yield the same result as lump sum taxes. For if the inelasticity results from the offsetting of income and substitution effects, a doubling of prices will generally have a different effect from the subtraction of half (or any other portion) of the individuals' original labor income, leaving prices unchanged. This is obviously so because the (alternative) new price lines will not have the same slope and so they cannot be tangent to the same indifference curve at the same point.

total revenue received by the firm or government: the fixed quantity of labor offered in exchange for their products. We have, for all price vectors in question, by the inelasticity of labor supply at quantity L_0:

$$\sum_j p_j x_j(p) = L_0;$$

$$\text{(7)} \qquad \text{i.e.,} \quad \sum_j p_j \frac{\partial x_j}{\partial p_i} + x_i = 0$$

for all i. Furthermore, as just noted, price changes can affect profits only through their effect on costs:

$$\text{(8)} \qquad \begin{aligned} \frac{\partial \pi}{\partial p_i} &= \frac{\partial}{\partial p_i}\left(\sum_j p_j x_j - C\right) \\ &= 0 - \sum_j \frac{\partial C}{\partial x_j}\frac{\partial x_j}{\partial p_i} \end{aligned}$$

Recall now the first form of "our" theorem which asserts that optimality requires, for all i,

$$-x_i = \lambda \frac{\partial \pi}{\partial p_i}$$

In the case now under consideration, this becomes by (8)

$$-x_i = -\lambda \sum_j \frac{\partial C}{\partial x_j}\frac{\partial x_j}{\partial p_i}$$

But by (7)

$$-x_i = \sum_j p_j \frac{\partial x_j}{\partial p_i}$$

Together these results imply that the first-order conditions can be written for all i,

$$\sum_j \left(p_j + \lambda \frac{\partial C}{\partial x_j}\right)\frac{\partial x_j}{\partial p_i} = 0$$

Since $\partial C/\partial x_j$ is the marginal cost of x_j, this asserts that prices bearing the proportion $-\lambda$ to marginal cost satisfy the first-order conditions.

In the more general case where shifting takes place both between sectors ($\partial C/\partial p_i \neq 0$) and within the taxed sector ($\partial x_i/\partial p_j \neq 0$, at least sometimes), an analogous algebraic chain of steps leads to the following first-order conditions, for all i:

$$\text{(9)} \qquad \sum_j \left(p_j + \lambda \frac{\partial C}{\partial x_j}\right)\frac{\partial x_j}{\partial p_i} = (1+\lambda)\frac{\partial L}{\partial p_i}$$

Only by chance will these conditions be satisfied by prices proportional to marginal cost, or by prices deviating from proportionality by a simple function of own elasticity. In this we agree with Lerner. However, the theorems prescribe definite prices which do optimize, given the constraints, and there is no necessity to resort to a rough compromise between two polar rules. One rule prescribes for the polar situations as well as for intermediate ones. This does not mean that a more precise spelling out of Lerner's compromise would not yield the same (correct) conclusions: it means only that the task which remains is not to find the general rule. That is already accomplished by (9). Rather, it is only necessary to translate this rule into a form compatible with Professor Lerner's very illuminating intuitions.

VII. *The Case Where All Items are Taxable*

Finally, we must yet deal with the case, excluded from our discussion thus far, in which "*all* uses of resources" can be taxed. In view of our preceding agreement with Lerner that when tax changes result in resource shifts only within the taxed sector, prices proportional to marginal costs are optimal, our assertion that this is generally not the case when all goods are taxable may come as a surprise. The explanation is, however, relatively simple: overall marginal cost pricing will not, as a rule, yield the necessary revenue.[14] This

[14] We shall say that "overall marginal cost pricing" prevails when the price vector confronting consumers is a scalar multiple of the vector of first partial derivatives of the transformation function evaluated at the associated social total net output vector.

follows from a basic assumption of the entire analysis: that lump sum taxes are impossible. In terms of our general equilibrium this requires (by Walras law) that the net value of consumers' sales and purchases of all commodities, *including labor*, add up to zero. In that case, a proportionate tax, leading to a proportionate change in prices, will still leave the net consumer expenditures at zero and hence will yield no revenue to the government.

In formal terms, let P_i be the price of commodity i $(i=1, \ldots, n)$ facing the consumer, no longer expressed in labor terms, and let P_L be the price of labor. Following the sign convention that positive numbers represent net purchases of desired goods, negative numbers, net sales, let $a_1, \ldots, a_n, a_L$, be the vector of net transactions carried out by a consumer. Our no-lump sum tax assumption can be expressed as the requirement that $P_1a_1 + P_2a_2 + \ldots + P_na_n + P_La_L = 0$. (This implies, of course, that the aggregate net transactions vector, summed over all consumers, satisfies the same condition, expressed above as the condition on the demand functions $p_1x_1 + \ldots + p_nx_n = L$, where prices were expressed in labor terms.) Obviously, here it makes absolutely no difference to the consumer whether he faces prices $P_1, \ldots, P_n, P_L$ or prices ten times as high; only relative prices count. If the government were to collect the difference between these two price vectors (P and $10P$) as a tax, it would collect precisely $9(\sum_i P_ia_i + P_La_L)$, which is zero. A tax vector which is a scalar multiple of the price vector facing consumers will, under our zero-transfer assumption, *always* yield zero revenue.

With consumer budgets equated to zero, the vector of taxes thus cannot be proportional to consumer prices if there is to be any tax yield, and hence consumer prices cannot be proportional to marginal costs. Now we are back where we started. If the government wishes to obtain, say, a quantity M of labor for its own uses, and can only work with commodity taxes, it must get consumers to settle at an equilibrium with an aggregate net transaction vector $x_1, \ldots, x_n, -L$, satisfying $L - F(x_1, \ldots, x_n) = M$, i.e., such that the leisure given up for production exceeds that necessary to produce the other desired goods by M.

VIII. *An Analysis in Terms of Quantity Changes*

Some readers may find it instructive to see a derivation of the propositions carried out explicitly in a general equilibrium context and in terms of quantity changes rather than price changes. This is easily done within our present model if it is assumed that lump sum redistributions maintain a socially optimal distribution of alternative social totals of goods and leisure. Then we can employ a Samuelsonian social utility function relating social welfare to these totals, with the property that the tangent plane to any social indifference surface is also a "budget plane" associated with prices at which that point will be sustained as a competitive equilibrium. Thus prices can be written as a function of quantities. The analysis is made more transparent if we assume that there is a certain fixed total, $\bar{L}$, of labor time which must be allocated among production, leisure, R (for recreation), and government surplus, yielding the constraint:

$$F(x_1, \ldots, x_n) + R + M = \bar{L}$$

where $F(x_1, \cdots, x_n)$ is obviously the (labor) cost function for the output bundle $(x_1, \cdots, x_n)$.

The problem is to maximize over output-leisure bundles the social utility function, $U(x_1, \cdots, x_n, R)$, subject to this constraint, and subject to the constraint that exactly the required government surplus is obtained as profit:

$$\sum_{i=1}^{n} p_i x_i - F(x_1, \ldots, x_n) = M,$$

where the p's are equilibrium labor-prices associated with the chosen output-leisure bundle. Associating the Lagrange multiplier λ with the first constraint, and the multiplier, μ, with the second, we can write out the Lagrangian expression:

$$\begin{aligned} V(x_1, \ldots, x_n, R, \lambda, \mu) &= U(x_1, \ldots, x_n, R) \\ &\quad + \lambda[\bar{L} - F(x_1, \ldots, x_n) - R - M] \\ &\quad + \mu\left[M - \sum_{i=1}^{n} p_i x_i + F(x_1, \ldots, x_n)\right] \end{aligned}$$

Taking partial derivatives we obtain the first-order conditions:

$$\frac{\partial U}{\partial x_i} - \lambda \frac{\partial F}{\partial x_i} - \mu\left[\frac{\partial}{\partial x_i}\left(\sum_{j=1}^{n} p_j x_j\right) - \frac{\partial F}{\partial x_i}\right] = 0, \quad i = 1, \ldots, n$$

$$\frac{\partial U}{\partial R} - \lambda = 0$$

Dividing each of the first n conditions by the $(n+1)$st and noting that equilibrium of the consumer requires $\partial U/\partial x_i/\partial U/\partial R = p_i/p_r = p_i$ (since labor is the numeraire, its price unity) we have

$$p_i - \frac{\partial F}{\partial x_i} = \frac{\mu}{\lambda}\left[\frac{\partial}{\partial x_i}\left(\sum_{j=1}^{n} p_j x_j\right) - \frac{\partial F}{\partial x_i}\right].$$

Employing, with their usual meanings, the symbols MC and MR (the latter including the effect on revenue of variations in *all* prices), the preceding condition implies

$$p_i - MC_i = \frac{\mu}{\lambda}(MR_i - MC_i), \quad i = 1, \ldots, n$$

which is the second form of our theorem from which the other forms can be obtained exactly as before. Thus we have derived the results explicitly from a general equilibrium model. Note that the term "marginal revenue" is definable here because we have normalized prices to labor dimensions and have assumed behind-the-scenes income distribution, making determinate the effect on equilibrium prices of changes in any of the desired outputs. Note also that under these conditions, the second version of the theorem, proved in Section III for the case of independent demands, is shown to hold generally.

IX. *Notes on the History of the Discussion*

There is some point in going briefly into the antecedents of the discussion because this history makes it all the more difficult to understand why the propositions in question have achieved so little recognition by the profession. The general line of argument has appeared widely for the better part of a century. The formal theorems themselves date back more than forty years. As we will see, this work has appeared in some of our leading journals under the authorship of some of the luminaries of our profession and was clearly not limited to a backwater of the literature.

There is an informal proposition very closely related to the results under examination that has a long history going back at least to the 1870's. With the establishment of the Railway and Canal Commission in England in 1873, and of the discussion preceding the establishment of the Interstate Commerce Commission in the subsequent decade there arose a rich literature examining utility pricing in relation to the public interest. A number of authors advocated prices that vary directly with demand *in*elasticity ("value of service").[15]

[15] For some examples see A. T. Hadley (ch. 6); E. Porter Alexander (esp. pp. 2–5, 10–11) and W. M. Acworth (ch. 3, esp. pp. 57–60). Hadley, the American

The mainstay of the early discussion was an argument only loosely related to the formal propositions in which we are interested. In brief, it was maintained that a relatively low price in elastic markets, provided it covers more than incremental cost, may well permit lower prices in other markets and hence be beneficial to everyone. For (particularly if the firm is subject to a constraint on its overall profit) the opening of a market which makes any net contribution may permit or may even require a reduction in prices elsewhere.[16]

Discussion along these lines has since appeared throughout the literature of public utility economics. (See, e.g., James Bonbright, esp. chs. 5, 17–20.) The various forms of the proposition have also appeared at least implicitly in a number of the standard writings of economic theory, notably in Joan Robinson's work and that of W. Arthur Lewis.[17]

However, the formal mathematical propositions which are derived from optimality considerations and which are the subject of this paper, themselves have a substantial history. As propositions on optimal taxation they first appear in 1927 in Frank Ramsey's pathbreaking article on taxation.[18] Thus, more than a decade before the publication of Hotelling's historic discussion of marginal cost pricing, Ramsey had in another context provided at least implicitly a solution to the optimal pricing problem for an industry in which marginal cost prices do not cover total costs.

A. C. Pigou took up the Ramsey discussion in the following year in his book on public finance, providing a very lucid and rather extensive summary of the argument. It was independently approached by Ursula Hicks in her 1947 book on the same subject. Her analysis is largely diagrammatic and is based entirely on the Marshallian consumers' surplus, just as Ramsey's and Pigou's had been. However, as far as we have been able to tell, the formal anal-

economist who went on to become a president of Yale, is the author of what is apparently the classic discussion of the period, in which he coined the phrase "not charging what the traffic can *not* bear." There is in fact a hint of the argument much earlier than this in the work of Dupuit who advocates the separation of consumers into different classes each paying a different price, remarking "suppose that (railroads) offered nothing but first class tickets, what a loss for the public and for the companies" (pp. 124–27).

[16] "If the (New York to San Francisco) price is more than the *additional outlay* involved in doing it, as against leaving it alone, it is profitable to the railroad, and the business is moreover advantageous to the whole inland community served by the railroad. For it adds to the number of men employed along the line and . . . the more prosperous the road, the lower the local rates may be made" E. P. Alexander (p. 4) (author's italics). The author adds (p. 5) that between the limits of "unreasonable" profits and net loss, actual profit rates "should be adjusted in proportion to value of service rendered" (*in*-elasticity of demand).

[17] See Joan Robinson (p. 207). Since she is dealing with a single product and does not utilize any explicit revenue constraint, the result she gives is not quite what is needed for our purposes but it is closely related to the theorem under discussion.

Lewis' statement (pp. 20–21), on the other hand, is precisely on target:

> The principle . . . is that those who cannot escape must make the largest contribution to indivisible cost, and those to whom the commodity does not matter much . . . get off lightly. . . . When there are escapable indivisible expenses to be covered the case for discrimination is clear. It secures an output nearer the optimum and levies the indivisible cost on those who get the greatest benefit (measured by their consumers' surplus). . . . Moreover, it is possible in some cases that the net result may be that everyone pays a lower price. . . . *If the undertaking is out merely to cover its costs* . . . reducing the price to some persons with elastic demands may increase the surplus over marginal cost which they contribute. . . . (emphasis added).

[18] The theorem seems to have made its greatest impact on the literature of public finance. A. C. Pigou (1928, pp. 126–28); (1947, pp. 105–09) reports Ramsey's results, offers an intuitive explanation and translates the theorem into elasticity terms. He also argues that except in the case of linear supply and demand functions the argument holds only if the taxes are small.

Ursula Hicks (pp. 167 ff) provides a very similar result on the basis of an elementary graphic argument. Yet even Richard Musgrave in his classic volume (p. 148) devotes only a footnote to the theorem and dismisses it as being "arrived at within the framework of the old welfare economics of interpersonal utility comparison" His characterization of the argument as part of the ability-to-pay approach is surely rather misleading. Cf. R. Bishop's comment on Musgrave, p. 212n.

ysis did not reappear until 1951 in an article by M. Boiteux which was the first to dispense with the notion of consumers' surplus. Apparently in the same year, Paul Samuelson submitted to the U.S. Treasury a closely related paper containing a generalization of Ramsey's results.[19] Samuelson's approach differs from that which we are discussing in that he employs no explicit revenue requirement. Instead he requires some preselected quantity of each output to be left unused by the private sector and made available to the government. His result, however, is essentially the same as those of the other contributors under discussion.

In 1952 there appeared a particularly simple derivation of the theorem in its most elementary form in a note by Alan Manne. His discussion assumes that cross-elasticities of demand are all zero and it measures and aggregates consumers' and producers' surplus in an unabashed Marshallian manner, but his derivation compensates for these simplifications by its great lucidity. Boiteux and Manne were also the first to present the analysis in the form of a discussion of public utility pricing. One year later Marcus Fleming provided an independent and illuminating derivation.

Meanwhile, Gerard Debreu had contributed analytic materials which led Boiteux to employ a Pareto optimality approach to the matter. On this basis Boiteux was then able to complete his definitive analysis of the subject. In Boiteux' classic 1956 article the analysis is, as a result, independent of any interpersonal comparison. The notions of consumers' and producers' surplus are dispensed with altogether. He is able to deal with goods whose demands are interdependent, with the pricing of inputs and intermediate goods as well as with outputs. It has also been emphasized by a subsequent writer (Jacques Drèze) that by his general equilibrium approach to the matter, Boiteux was contributing an explicit second-best solution, one of the few to be found to date in any area of welfare economics. Of course, no piece of analysis can answer every relevant question, and Boiteux left unsettled some matters of detail. There is a moderately mysterious role played in his analysis by the numéraire commodity; he has not dealt with problems of nonnegativity of outputs, etc. All this has left room for subsequent work by R. Rees, William Vickrey (*Testimony* and other unpublished work), the present authors, and by a number of others.[20]

[19] Samuelson's own brief description of the history of the analysis (1964) is worth reproducing:

> . . . Consider an optimal *laisses faire* situation that maximizes a social welfare function with zero government expenditure and taxes. Now introduce government services as a (vector) function of a small or large parameter y, $G(y)$. Suppose excise taxes (on goods or services) to be alone feasible, and introduce a (vector) pattern of excises $T(y)$ sufficient to provide resources for the G. What is the optimal $T(y)$ pattern to maximize the welfare function $W(y)=W(T, G)$? This is the problem set by Pigou to Frank Ramsey in the 1920's. Approximate answers were given by Ramsey, Boiteux (and in unpublished form by Hotelling and Hicks). In a pearl cast before the U.S. Treasury in 1950, I gave an exact solution for large and small programs; namely, that the optimal pattern is the one at which the response of all goods and factors to a further compensated-Slutsky price distortion would result in equal percentage (virtual) reductions.
>
> My literary wording is loose, but there is no need for approximative consumer's surplus at all. . . .

Though the reported date of the Samuelson memorandum may seem to imply some ambiguity about the order of appearance of this and the first Boiteux article, this issue is cleared up by Samuelson's explicit reference to the Boiteux article on page 5 of his memorandum in which he expresses his debt to the "brilliant analysis by M. Boiteux."

[20] Among those who have also written on the subject are Diamond and Mirrlees and Bishop. Recall in this connection Samuelson's reference to unpublished work by Hicks and Hotelling, on which more light is shed by the following excerpt from a letter which the authors received from Professor Hicks:

> It was just as well that you sent me the cutting from Samuelson [the preceding footnote], or I should have found it hard to recollect the matter. But with this refreshment it comes back. I think the story is as follows.

However, there is no need to go beyond Boiteux' work for a basis for the welfare discussion contained in this paper. The generalization to optimal pricing for an entire economy requires only one small modification in Boiteux' two-sector analysis. In his model there are two sets of firms, the first composed exclusively of perfect competitors, while the second are subject to the author's rules for quasi-optimal pricing. For our purpose we need merely take all firms to fall into the second category and our interpretation follows at once. In sum, it follows for the economy as a whole that unless marginal cost pricing happens to provide returns sufficient to meet the social (governmentally determined?) revenue requirement, a quasi-optimal allocation calls for systematic deviations of prices from marginal costs *throughout the economy* in the manner specified by the theorems that this paper has described.

X. *Concluding Comment*

We conclude our survey of the relevant literature as we began—at a loss to explain how a set of results flowing from the pens of so distinguished a set of authors and appearing and reappearing in the profession's leading journals, should so long have escaped general notice. This is all the more curious in light of the importance of the subject and the elegance of the theory.

As a theoretical matter, the theorem we have discussed seems fundamental not only for the analysis of public finance and the regulation of utilities, but also for some of the basic precepts of welfare economics. Earlier we spoke of the result as a proposition in the theory of the second best. But it is more than this. In a world in which marginal cost pricing without excise or income taxes is normally not feasible, the solution we have usually considered to characterize the "best" is none too good, because it is simply unattainable. In that case, the systematic deviations between prices and marginal costs that the theorem calls for may truly be optimal because they constitute the best we can do within the limitations imposed by normal economic circumstances.

In December 1946 (in the last week of my first visit to the United States) I stayed for a couple of days with Hotelling at Chapel Hill. I then told him that I had worked out (using a quite crude consumers' surplus method) that if a given sum had to be raised in excise taxes, the least-sacrifice way of doing it was the proportional all-round reduction in consumption, as compared with an optimum position. For if the burden of the tax is

$$\frac{1}{2}\sum q_{ij}dp_i dp_j$$

this is minimised subject to $G=\Sigma q_i dp_i$ if

$$dq_i = \sum q_{ij}dp_j = \lambda q_i$$

where λ is the Lagrange multiplier. Not much more than this, except that I had allowed (in the spirit of the old Hotelling article) that the q_{ij} could incorporate supply as well as demand reactions.

I remember he told me I ought to publish this, but I didn't—mainly, I suppose, because I was conscious of the qualifications to which Samuelson alludes in his paper, and which, if I had set them out in my style, would have whittled away the result so near to nothing.

Appendix

Comments On Lerner's Appendix

In the Appendix to his paper, Lerner asserts that our analysis over-emphasized demand elasticities. In addition, the Appendix contains a wider range of lesser criticisms. To deal in depth with all of these would require us to impose on the reader's time and patience. We propose to touch briefly on what appear to us to be the more important issues raised by Lerner, and trust that our failure to respond to other points will not be taken as a sign of general agreement.

Two criticisms appear recurringly in Lerner's Appendix and they are best disposed of at the outset. The first is that we have somehow used as a standard of welfare a measure of *consumers' loss*, and have left out any measure of *social damage*. Since our economy was taken to consist entirely of consumers, i.e., we use the terms "consumer" and "person with utility function" here

interchangeably, we would use the two terms "consumers' loss" and "social damage" interchangeably as well. We would use neither without a grain of salt. The touchstone of quasi-optimality is plain and simple Pareto optimality, given the profit or tax revenue constraint. As is by now generally recognized, Pareto optimality cannot be comfortably equated with a "higher ethical good," and we apologize if we give the impression that a quasi-optimal price structure (for the monopolist) or tax structure was "good" in any sense other than Pareto optimality.

The second assertion that occurs in several places is that our analysis assumes constant marginal costs. This is simply incorrect. The only point at which marginal costs are taken to be constant is in our Section IV, the graphical presentation, which is intended solely as an aid to intuition. Elsewhere there is no assumption of any sort about marginal costs and none of the results depend on such an assumption. Lerner seems to suggest that our assumption that the prices of inputs are fixed can be equated to the premise that marginal cost is constant. We fail to see any connection between the two. Our assumption that input prices are constant is another, perhaps mistaken, compromise in the direction of intuitive appeal.[21]

Incidentally, the point of our footnote 7 is not, as Lerner suggests, to acknowledge in passing that marginal costs might in fact be variable. Rather, it is intended to point out that the marginal costs in question in the various forms of the theorem are *not* those which prevail in some initial situation before any taxes are imposed, but those which in fact rule at the solution values. Rather than apologizing for assuming marginal costs constant, we are, on the contrary, saying that marginal costs will very likely change as taxes are varied, so one must watch out and be sure to use the correct values in testing for optimality.

The basic task undertaken by Lerner in his Appendix relevant to the discussion here is derivation of our four forms of the "mislaid maxim" from his second rule. Since all four are derivable from either the first ($\partial\pi/\partial p_i=\lambda x_i$, all i) or the fourth ($\Sigma_j(\partial x_i/\partial p_j)(p_j-MC_j)=\gamma x_i$, all i) it is reasonable that he should be able to do this. It is nevertheless somewhat surprising that he did not in the process notice that his second rule takes into account only "own tax elasticity" and hence implicitly assumes the zero cross elasticities which he finds so objectionable in our second and third forms. It is probably because of this that he comes to the conclusion that our first form holds only when cross-elasticities are zero, which is simply incorrect.

Lerner's handling of our first form of the theorem (which, incidentally, is the one form we did *not* distill from the literature although since deriving it, we have seen essentially the same variant in Diamond and Mirrlees) is altogether curious. Almost paradigmatic of the economist's approach to a welfare problem, say, how large a park to build, is the following: "Clearly our rule should be, enlarge the park until the incremental social gain from enlargement by an additional unit is just balanced by the social loss from the space and other resources thereby foregone. But that is, of course, purely tautological. What we now need to do is clothe the concepts of social gain and social loss with operational measures. . . ." In this case Lerner goes backwards: "The equalization of the ratio between marginal yield and output level turns out to be an obscure way of expressing the equalization of marginal tax yield and marginal social damage—an equalization that minimizes the total social damage for a given total tax yield" (p. 289). We were, as it happens, especially pleased with the first form of the theorem which we think is new, precisely

[21] Analytically, it does serve to a set a price level so that the government profit constraint in pecuniary form is pinned down. In the case of more than one input factor it does, however, hide rather strong assumptions about the transformation function in a full general equilibrium treatment. In any event, the pecuniary variant of the government's budget constraint is perhaps not the most felicitous for use in such a treatment. The theorems in question are, however, in no way sensitive to these assumptions. Again we recommend perusal of the literature on this point.

because it seemed to promise more than the others a degree of operational applicability. Our feeling was that experienced tax officials (or monopoly managers) would have a much better chance of estimating at least roughly the magnitudes that enter that expression—the effect on the total tax take (profit) (including cross effects) of a change in a given tax (price) than econometricians would have of estimating the full matrix of cross elasticities, just as it seems plausible that a businessman may come close to maximizing profits even though he is pretty vague about quantities such as marginal cost and marginal revenue. If this is misplaced faith it doesn't matter, because the first form of the theorem, being equivalent to the others (and for more forms see our forthcoming paper), is at least no harder than they to put into practice. One thing is clear, the operational meaning of the first form is easier than the others (except, perhaps, for the fourth) to explain to a non-economist.

The derivations of the second and third forms of our theorem, like those of the first and fourth, make no use of and in no way depend upon constancy of marginal cost. The second and third variants do, however, involve an assumption of zero cross elasticities of demand among the goods whose prices are in question. Lerner is quite properly disturbed by the implication of extending this assumption to *all* goods. The demand relationship being traced out by variations in the price vector is simply the aggregate offer curve of the economy with the initial commodity vector of each citizen specified (generally at zero) and no net transfers. Independent demands for the individual goods are impossible here except in the trivial case in which the demanded vector is completely independent of the price vector, a vector which would in turn have to be the zero net transactions vector. The implication is not that the results are incorrect but that the second and third forms are not relevant when all goods are taxable and there are no net transfers.

Again, contrary to Lerner, the fourth form of the theorem is as general as the first, and more general than the second and third (*not* because it doesn't depend upon constant costs—they don't either—but because it does not assume zero cross-elasticities). Given the parenthetical word "approximately" its generality would seem assured. But we were willing to go farther than that, and took great pains to make precise in Section III, in the algebraic development of the fourth variant, what operationally was meant by "approximately" here. In footnote 9 we spell this out still further.

REFERENCES

W. M. Acworth, *The Railways and the Traders.* London 1891.

E. P. Alexander, *Railway Practice.* New York 1887.

R. Bishop, "The Effects of Specific and ad valorem Taxes," *Quart. J. Econ.*, May 1968, *82*, 198-218.

D. F. Bradford and W. J. Baumol, "Quasi-optimal Pricing with Regulated Profit and Interdependent Demands," forthcoming.

M. Boiteux, "Le 'revenue distribuable' et les pertes économiques," *Econometrica*, Apr. 1951, *19*, 112-33.

———, "Sur la gestion des Monopoles Publics astreints à l'équilibre budgétaire," *Econometrica*, Jan. 1956, *24*, 22-40.

J. Bonbright, *Principles of Public Utility Rates.* New York 1961.

G. Debreu, "The Classical Tax-Subsidy Problem," *Econometrica*, Jan. 1952, *20*, 100.

P. A. Diamond and J. A. Mirrlees, "Optimal Taxation and Public Production," mimeo. working paper 22, M.I.T. 1968.

J. H. Drèze, "Some Postwar Contributions of French Economists to Theory and Public Policy, with Special Emphasis on Problems of Resource Allocation," *Amer. Econ. Rev.*, June 1964, *54*, 1-64.

J. Dupuit, *Traité théorique et pratique de la conduite et de la distribution des eaux.* Paris 1854.

J. M. Fleming, "Optimal Production with Fixed Profits," *Economica*, Aug. 1953, N.S. *20*, 215-36.

A. T. Hadley, *Railroad Transportation.* New York and London 1886.

J. R. Hicks, *Value and Capital*, 2d ed, New York 1956.

Ursala Hicks, *Public Finance.* New York 1947.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica,* July 1938, *6,* 242-69.

H. Leibenstein, "Allocative Efficiency vs. 'X-Efficiency,'" *Amer. Econ. Rev.,* June 1966, *56,* 392-415.

A. P. Lerner, "On Optimal Taxes With an Untaxable Sector," *Amer. Econ. Rev.,* June 1970, *60,* 284-94.

W. A. Lewis, *Overhead Costs.* London 1949.

A. Manne, "Multiple-Purpose Public Enterprises—Criteria for Pricing," *Economica,* Aug. 1952, N.S. *19,* 322-26.

R. Musgrave, *The Theory of Public Finance.* New York 1957.

A. C. Pigou, *A Study of Public Finance,* (a) 1st ed., London 1928; (b) 3d ed., London 1947.

F. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.,* Mar. 1927, *37,* 47-61.

R. Rees, "Second-Best Rules for Public Enterprise Pricing," *Economica,* Aug. 1968, N.S. *35,* 260-73.

Joan Robinson, *The Economics of Imperfect Competition.* London 1933.

P. A. Samuelson, "Theory of Optimal Taxation," unpublished, approx. 1951.

———, "Principles of Efficiency—Discussion," *Amer. Econ. Rev. Proc.,* May 1964, *54,* 94-95.

W. Vickrey, *Testimony,* F. C. C. Docket No. 16258, Networks Exhibit No. 5, Appendix I. Washington 1968.

O. Williamson, *The Economics of Discretionary Behavior.* Englewood Cliffs 1964.

On Optimal Taxes With an Untaxable Sector

By ABBA P. LERNER*

In the absence of externalities the optimal allocation of resources between different uses is reached by setting the prices of products at marginal (opportunity) cost—the price of the alternative product—so that the consumers' demand will cause that alternative to be produced which they prefer to consume.

I

If for any reason more money must be collected from the public, the optimal allocation of resources will be preserved if the money is raised by "ideal" taxes, namely taxes which do not entail social damage by inducing resources to be shifted to less urgently desired purposes for the sake of avoiding the taxes. Such ideal and "unavoidable" taxes are (1) a "lump sum" tax and (2) a uniform tax on all uses of resources. A lump sum tax makes shifting irrelevant, and a uniform and universal tax cannot be avoided by shifting since the same tax is waiting wherever the resources are shifted. Lump sum taxes being impractical or perhaps impossible, this gives us *the first rule* of ideal taxation: Maintain the *proportionality* of price to marginal cost by having a uniform *tax rate* on all uses of resources by means of an ad valorem tax on *all* final output. *Equalize* T/MC, the ratio of the *tax* to the *marginal cost*. This would *eliminate* or prevent any social damage and the outcome would be Pareto optimal.

But it is no more possible to tax *all* final outputs than to have lump sum taxes. Some uses of resources, like leisure, do it yourself, or undeclared mutual services, are *untaxable*, and it will not do to tax only *some* of the outputs proportionately. We are therefore left with the question: How should *taxable* uses be taxed?[1]

The first category to be chosen for taxation consists of items with completely inelastic demand or completely inelastic supply. This works like a lump sum tax since there is no elasticity of output with respect to the tax and no shifting at all.

If enough money cannot be raised in this way (because as tax rates are increased, output must be affected—i.e., some "tax elasticity of output" emerges), taxes will have to be imposed where there is some tax

* University of California, Berkeley. This paper was inspired by the article by Professors William J. Baumol and David F. Bradford in this issue, and gained very much from correspondence with them. It also benefited from discussion with my colleagues George Akerlof and Avinash Dixit, and would have benefited still more if I had listened to them more carefully.

[1] The treatment of this question constituted a significant part of my contribution to the "Lange-Lerner" socialist economics. It took the form of rejecting a proportional *negative* tax (social dividend). "Dr. Lange declares that the social dividend must be distributed as a percentage on the wage rate if it is not to interfere with the ideal distribution of labour between different occupations (p. 21). This seems to me to be an error . . . If the social dividend is made proportional to the wage, there will be an undue attraction of workers to the occupation with the greater wage on account of the greater social dividend obtainable there in addition to the greater wage" (Lerner, 1936 p. 73). The issue is also dealt with in *The Economics of Control*, ch. 9 "Equality and Proportionality" pp. 96–105, where the headlines read: "Making price *proportional* instead of equal to marginal cost was believed to be enough . . . but proportionality cannot be universal unless it is really equality. . . . This is illustrated in the allocation of labor power between labor and leisure."

elasticity, and therefore some shifting of resources and some social loss. The taxation should then be concentrated where the tax elasticity is least since that is where the shifting is least.

T, the *tax* per unit, will be equal to $P-MC$, the excess of P, the *price* of the product, over MC, its *marginal cost*. But if P measures the marginal utility of the product to the consumer and MC measures the marginal disutility of producing it (basically the marginal utility of the alternative product foregone), then $P-MC$, or T, will also be the measure of MSD, the *marginal social damage* per unit reduction of this output and the consequent shifting of resources to the alternative use. If the total social loss (from a given amount of tax revenue to be collected) is to be minimized, the ratio between the extra damage due to, and the extra revenue collected by, a marginal tax increase must be the same everywhere. Otherwise a shift of taxation could increase tax revenue even while reducing the social damage. This gives us a *second rule* of ideal taxation: Adjust the tax rates on the different items so as to *equalize* T/MTR (where MTR is the *marginal tax revenue* from that tax increase which reduces output by one unit). This would not eliminate, or prevent, social damage but would *minimize* it and the outcome would be a "quasi-optimal" second-best solution of the problem.

The equalization of T/MTR implies the equalization of $T/(T-MTR)$. Since T, the tax per unit, is also ATR, the *average tax revenue*, this also implies the equalization of $ATR/(ATR-MTR)$. But average tax revenue over average *minus* marginal tax revenue is nothing but the elasticity of output with respect to the tax or the *tax elasticity*. The second rule may, therefore, also be expressed as: *Equalize the tax elasticities*. This form of the second rule may be seen as the result of concentrating the taxation where the tax elasticity is least, since any inequality of tax elasticity would be removed by shifting taxation from where the elasticity is greater to where it is less.

II

It must be recognized, however, that only if the shifting of resources is from the taxed sector to an untaxed sector of the economy does MC measure the usefulness of the alternative product, and only in that case is T or $(P-MC)$ equal to MSD. Equalizing tax elasticities (in equalizing T/MTR) thus minimizes only that part of the social damage which is due to the shifting of resources from the taxed sector to the untaxed sector. But equalizing the tax elasticities for different items with different demand elasticities and different supply elasticities involves *different tax rates*. These different tax rates will bring about a faulty allocation of resources between the different activities *within* the taxed sector. The loss from *this* misallocation of resources would be not merely minimized but *eliminated* by following not the second rule but the first rule, i.e., by equalizing not T/MTR but T/MC.

Thus we see that the second rule is appropriate where the shifting is only from a taxed to an untaxed sector, while the first rule is appropriate where the shifting is only to another part of the taxed sector (as would be the case if all uses of resources could be, and were, taxed).

What we need is a rule that is applicable whether the resources shifted by a tax remain in a taxed sector or go to the untaxed sector or are divided between the two sectors. For this we can use a more general rule that encompasses both the first rule and the second rule: *Let the tax on each use of resources be equal to the tax on the alternative use.* There will then be no incentive to shift any resources to another (and inferior) use so as to avoid the tax. This would not merely minimize any social

damage from the taxation, it would *eliminate* it and the outcome would be Pareto optimal.

III

To be able to apply this more general rule, we would have to know in each case what the alternative use is and what taxes would fall on this alternative use. If the alternative use is entirely in the untaxable sector of the economy, the item in question should not be taxed either. It must be assimilated to the untaxable sector in a larger *untaxed* sector, reducing the taxable sector to a smaller *taxed* sector. On the other hand, if the alternative use is entirely in the taxed sector, the tax must be the same as in the alternative use. The two uses should be assimilated and considered as a single industry for our present purpose.

If there are only these two possibilities, then the required tax revenue should be raised by a uniform ad valorem tax on all sales in the taxed sector. Indeed, in such a case, a whole of the taxed sector can be considered as a single assimilated industry with absolutely inelastic demand or absolutely inelastic supply. No resources can be shifted out of it, and the tax is "unavoidable" and ideal, just like a lump sum tax.

But these two are not the only possibilities. There will be items from which only a *fraction* of the resources shifted will go to other parts of the taxed sector, the rest going to the untaxed sector. On such an item, if it were the only one of its kind, the ideal tax is this fraction of the otherwise *uniform* tax just mentioned, since that would be the amount of the tax on the alternative use of the resources. But the existence of such items means that there must exist sectors with various *lower* tax rates. The ideal tax for each good would, therefore, have to be adjusted further to allow for the different fractions of the shifted resources which would go to various sectors subject to different tax rates.[2,3]

IV

It would not seem possible to work out all these complications. They could be avoided if we could consider only the damage due to the shifting of resources from the taxed sector to the untaxed sector of the economy, neglecting the damage caused by the shifting of resources between the different parts of the taxed sector taxed at different rates. In that case we could simply apply *the second rule:* Equalize T/MTR. But the second rule is appropriate only if there is no substitutability, either in consumption or in production, between the different taxed items taxed at different rates, or if only a small sector of the economy is to be taxed (possibly because only a small sector is technically taxable) so that the shifting of resources among the different parts of the taxed sector can be considered small enough to be neglected, the bulk of the shifting being from the taxed to the untaxed sector of the economy.

The complications could also be avoided if we could do the opposite and consider only the social damage due to the shifting of resources between the different parts of the taxed sector, neglecting the social damage caused by the shifting of resources from the taxed to the untaxed sector. In that case we could simply apply *the first rule:* Equalize T/MC. But the first rule is appropriate only if there is no substitutability, either in consumption or in produc-

[2] My colleague Avinash Dixit has developed a mathematical formula which seems to correspond to such an application of the more general rule!

[3] It should be noted that one of the results of this would be that the tax rate would be lower the greater the degree to which it would cause resources to be shifted from the taxed to the untaxed sector, both directly and indirectly (i.e. via being shifted to other parts of the taxed sector which are being taxed at a lower rate because more of *their* resources are shiftable, directly or indirectly, to the untaxed sector, etc.).

tion, between the taxed and the untaxed sectors, or if the untaxed sector is very small so that the shifting of resources from the taxed to the untaxed sector can be considered small enough to be neglected. Where there is significant substitution both between the two sectors as well as between the different parts of the taxed sector, neither rule can be ignored and we cannot bypass the complications. A simplifying compromise must then be sought between the two conflicting rules.

V

Such a compromise could be based on the simplifying assumption that all resources shifted by a tax would be divided between the taxed and the untaxed sectors in proportion to their size; i.e., if the taxed sector is twice as big as the untaxed sector, twice as much of any shifted resources would go to other parts of the taxed sector as would go to the untaxed sector.

On this assumption the proportions of the revenue raised in accordance with the two different rules should be in proportion to the relative size of the taxed and the untaxed sectors of the economy. Thus if the taxed sector is twice as great as the untaxed sector, two-thirds of the required revenue should be raised by tax *A*, a uniform proportional tax on marginal cost, and one-third of it by a non-uniform tax *B*, concentrated where the tax *B* elasticities are least so as to equalize these for the different taxed items. (The imposition of each tax would of course reduce the revenue from the other tax, and if this had not been taken into account the tax rate would have to be recalculated.)

The combination of the two taxes, since it leaves the *B* tax elasticities equalized, looks something like the abandonment of our first rule in favor of the second. But this is not so. It is only the *B tax elasticities* that are equalized, not the overall tax elasticities. Only in the limiting case where

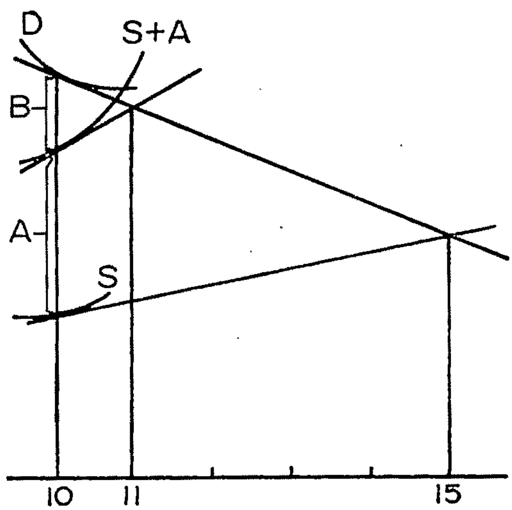


FIGURE 1

all shifted resources go to the untaxed sector does tax *A* disappear. The formula then degenerates into the second rule. In the converse limiting case, where none of the shifted resources would go to the un-

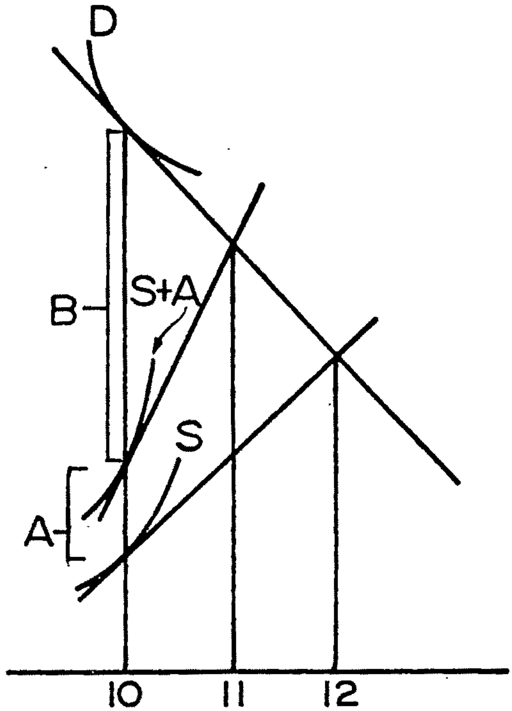


FIGURE 2

taxed sector, tax B disappears and the formula degenerates into the first rule.

Figures 1 and 2 show the tangents to the demand curves (D), to the supply curves (S) and to the supply-plus-tax-A curves ($S+A$), at the equilibrium output of 10 units. In both figures tax A is 100 percent on marginal cost (supply) and the B tax elasticities are 1/10 in each case. The overall tax elasticities, however, which would have been equalized by the application of the second rule, are 5/10 and 2/10, or $\frac{1}{2}$ and $\frac{1}{5}$, respectively.

APPENDIX

1. *"The Mislaid Maxim"*

In their article, "Optimal Departures from Marginal Cost Pricing," Baumol and Bradford (hereafter referred to as BB) survey the literature on the subject and rediscover in it a "mislaid maxim" which seems to me to overemphasize somewhat the departures from the principle of marginal cost pricing, as in their conclusion that "Generally prices which deviate in a systematic manner from marginal cost pricing will be required for an optimal allocation of resources." The mislaid maxim consists of a "basic theorem" which "has been stated in a number of different ways . . ." and which BB summarize in four different variants.

> 1) If prices are quasi-optimal the ratio between the marginal profit yields of unit changes in the *prices* of any two goods will be equal to the ratio between their output levels.
>
> 2) For each product, the deviation of the quasi-optimal price from marginal cost must be proportionate to the difference between the product's marginal cost and marginal revenue (i.e., its marginal social welfare cost must be proportionate to its marginal contribution to the profit requirement). This result holds only in the case where cross-elasticities are all zero.
>
> 3) For each product, the percentage deviation of quasi-optimal price from marginal cost must be inversely proportionate to its price elasticity of demand. This result also holds only where cross-elasticities of demand are zero.
>
> 4) Quasi-optimal prices must yield outputs that deviate by (approximately) the same proportion from those which would result from pricing at the marginal costs corresponding to the quasi-optimal output levels.

Of the four variants of the "mislaid maxim" the first, which requires "that the ratio between the marginal profit yields of unit changes in the prices of any two goods be equal to the ratio between their output levels," is derived by BB from a concern not with the *social damage* but with the *consumer's loss* from having to pay a higher price (which loss, for a unit increase in price, is identical with the output level). This loss, it should be noted, is still there even if the demand is absolutely inelastic, though the tax is then "unavoidable," equivalent to a lump sum tax, and causes no shifting of resources and *no social damage* at all.

Nevertheless, with the assumption of a constant MC or supply curve (which would be the case if we let the prices of all inputs supplied in the economy be fixed, and let competition eliminate economies and diseconomies of scale), it can easily be shown to follow from our second rule. If we so choose the (arbitrary) size of the units in which the output is measured as to make the tax rate per unit the same for both goods, equal *absolute* increases in P, will constitute equal *proportional* changes in T. With the tax elasticities the same (our second rule), *equal absolute* increases in P (since they constitute *equiproportional* increases in T) will be accompanied by *equiproportional* decreases in output. If the output of one of the goods was three times as great as that of the other before the increase in P, it will still be three times as great after the increase in P, and the yield in profit (or tax revenue) from a unit increase in price (or tax) would also be three times as great, thus satisfying the required proportionality between marginal yield and level of output.

This is shown in Figure 3 where OR measures the normalized T. The increase in revenue from the increase in price per unit

is shown in the horizontal rectangle RD, the loss in revenue from the reduction of output is shown by the vertical rectangle CD and the marginal profit yield from the price increase by the difference $RD-CD$. If the output of the other good, with the demand curve P^*, is three times as great, the corresponding rectangles will also be three times as great. RD^* has the same thickness as RD but is three times as long, and C^*D^* has the same height as CD but is three times as wide since the demand curve P^* is three times as flat. The marginal profit yield, $RD^*-C^*D^*$, will therefore also be three times as large and proportional to the three times larger output level.

The equalization of the ratio between the marginal tax yield and the output level seems to have no economic meaning. It is therefore somewhat strange that the formula should give the right answer. The explanation of the paradox is that an economically meaningful relationship is hidden within it.

The proportionality between output level (x) and marginal yield (y) derives from the proportionality between the areas of the rectangles RD and RD^*, which measure the consumers' loss (l) and the rectangles CD and C^*D^*, which measure the social damage (d). We thus have proportionality in *four* items, x, y, l, and d. The meaningful relationship, hidden behind the proportionality of x and y, is the proportionality of y and d. The equalization of the ratio between marginal yield and output level turns out to be an obscure way of expressing the equalization of marginal tax yield and marginal social damage—an equalization that minimizes the total social damage for a given total tax yield.

However, BB's first variant will be right only where there are constant costs and where all the shifting is into the untaxed sector of the economy. It can be generalized to cover increasing (or decreasing) marginal costs by substituting "taxes on" for "prices of," making it read: "that the ratio between the marginal profit yields of changes in the *taxes on* any two goods be equal to the ratio between their output levels." It then becomes equivalent to our second rule. But to deal with shifting within the taxed sector we have to turn to our first rule.

The second variant, which "calls for the deviation of [ ] price from marginal cost for each product to be proportionate to the difference between its marginal cost and its marginal revenue" is identical with the first form of our second rule (equalize T/MTR) where $T=(P-MC)$ and $MTR=(MR-MC)$ (MR [*marginal revenue*] including the marginal tax and MC excluding it).

The third variant requires that "for each product the percentage deviation of [ ] price from marginal cost must be inversely proportional to its price elasticity of demand." Like the first, it can be derived from our second rule in combination with the assumption of constant marginal cost. If the marginal cost is constant, our second rule (Equalize the tax elasticities) is equivalent to making the tax rate inversely proportional to the elasticity of demand. This is shown in Figure 3. Taking the horizontal MC curve beginning at 0 as the base, the height of the P (or demand) curve above the MC (or supply) curve measures T (the tax per unit). At point D this will be equal to CD. The tangent of the demand curve at D (AB'') shows us that

$$e_T \text{ (the tax elasticity)} = BD/DA$$

If the base line from which P (the price, including the tax) is measured is OB' or $O''B''$, for two different goods, then

$$e_D \text{ (the elasticity of demand)} = B'D/DA \quad \text{or} \quad B''D/DA$$

$$T/P \text{ (the \textit{tax rate})} = \frac{CD}{C'D} \text{ or } \frac{CD}{C''D} = \frac{BD}{B'D} \text{ or } \frac{BD}{B''D} = \frac{BD/DA}{B'D/DA} \text{ or } \frac{BD/DA}{B''D/DA} = \frac{e_T}{e_D}$$

Since e_T is a constant, the same for all the taxed industries, T/P, the tax rate, is inversely proportional to the elasticity of de-

mand for the different products. (This, BB's third variant, must however be measured as a percentage of the *price* and not of the *marginal cost*—it is T/P not $T/MC[=T/(P-T)]$ that must be proportional to the elasticity of demand.) Here they pay their respects to marginal costs being "not generally constant" and point out that Ramsay, Pigou and Boiteux did take *elasticity of supply* into account.

The arguments supporting the second and third variants make use of the "simplifying assumption" that the cross-elasticities are all equal to zero. If this is taken strictly, it means that there is no shifting at all of resources to alternative (and less preferred) untaxed or more lightly taxed products. The resources set free are just those required to satisfy the requirements of the government (or there may be no resources set free at all if the purpose of the taxation was only to prevent overall excess demand inflation). It also means that the demand elasticities will all be *equal* (at unity, since if there is no change in expenditure on any other item, there cannot, with the same total income, be any change in the expenditure on "this" item). All the variants of the second rule which say the tax must be proportional to the elasticity of demand then become indistinguishable from our first rule which says that the tax must be *proportional*. Period. (The elasticity of supply has been eliminated from the picture by the assumption of constant costs.) But the situation is stranger still. *Any* tax system whatsoever is now optimal since there can in any case be no shifting and no social damage from the switching to less preferred goods for the sake of avoiding taxation.

If the assumption of zero cross-elasticities is taken to refer only to the taxed items, then any shifting of resources *within* the taxed sector is assumed away and there remains only the possibility of the shifting of resources from the taxed to the untaxed sectors of the economy. The second rule, and any variant of the maxim which implies it, then minimizes the damage from this shifting, and is quasi-optimal.

The fourth variant ". . . [ ] prices must yield outputs that deviate by (approximately) the same proportion from those which would result from pricing at the marginal costs . . . ," is more general than the first and third, as is claimed by BB, in that it does not depend on constant costs. Like the other three variants it fails to deal with the social damage from tax-induced shifting of resources within the taxed sector. It is *less* general than the other variants in that it will be true only if the supply and demand curves are straight lines (or are of exactly offsetting opposite convexities) so that the tax revenue curves (T' in Figure 3) obtained by subtracting the supply curves (vertically) from the demand curves (which would show the output corresponding to each level of tax per unit of output) are straight lines. However, if these curves are approximately straight, the result will be approximated.

2. *Constant Costs*

The assumption of constant cost, by making the elasticity of supply infinite, is responsible, in the first and third variants and in much of the literature, for an undue concentration on the elasticity of demand and for obscuring the completely symmetrical relevance of the elasticity of supply, or of the MC.[4] Thus, if the supply (or MC) curve is less than infinitely elastic (represented in Figure 3 by MC') and the demand curve is infinitely elastic (represented by the horizontal line OC^*), the tax (EC) that equalizes the tax elasticity constitutes a *tax rate* (here represented as T/MC and seen as EC/EC' or EC/EC'') inversely proportional to the elasticity of *supply* instead of to the elasticity of demand. If neither the supply curve nor the demand curve is infinitely elastic (and they are represented by P and MC' respectively) the tax (ED or SR) that equalizes the tax elasticities constitutes a *tax rate* proportional to a *combination* of the elasticity of supply and the elasticity of demand. It is equivalent to a tax rate on the buyer of $CD/C'D$ or $CD/C''D$, i.e., proportional to the inverse of the elasticity of demand, *plus* a tax rate on the seller of $EC/$

[4] This is what enables BB to be satisfied with the four variants of the second rule.

$C'E$ or $EC/C''E$, i.e., proportional to the inverse of the elasticity of supply.[5]

The assumption of constant costs is not merely a slip or a simplification. It is the result of a widespread confusion of (a) the distinction between *demand* and *supply* with (b) the distinction between *preferences* (among outputs or inputs) and the *technically given production possibilities*. The *sound* consideration that social damage from tax-avoiding shifting of resources can consist only in faulty allocation in relation to *preferences* (if only final products are taxed) is thus translated into the *unsound* conclusion that only the elasticity of *demand* is relevant.[6] This can be seen in the figure used by BB in their article where the possibility of any producers' surplus is eliminated by the assumption of constant cost. The rectangles (CK and CE_b) between the horizontal marginal cost curve CC' and the demand curve D_b represent the transfer of income from the consumer to the monopolist as monopoly revenue (or to the government as taxes) at prices P_0 and P_1. The increase in this transfer from an increase in P (or T), namely CE_b *minus* CK or P_0E_b minus R_bK, is the increase in net (monopoly or tax) revenue. Subtracting from this the part (P_0E_b) which consists of the higher price paid by the consumer on the reduced output —approximately his loss of consumers' surplus—we are left with the *minus* item (R_bK) and the conclusion that "the rise in price will have caused a net reduction in net revenue *plus* consumer's surplus that is measured by the shaded area R_bRKE_b."

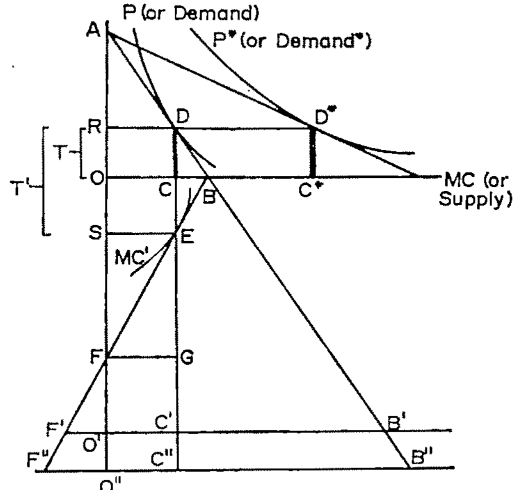


FIGURE 3

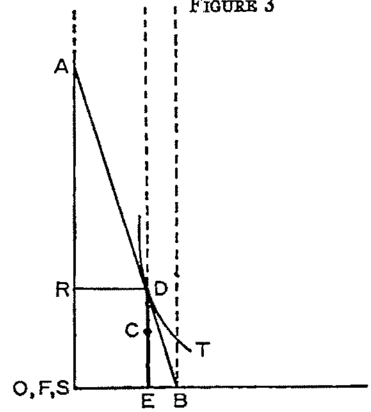


FIGURE 4

This conclusion is not incorrect, but it throws a very dim light on the subject. The shaded area R_bRKE_b does measure the degree to which the loss of consumer's surplus is greater than the gain in revenue, but it is not instantaneously clear that this difference is the measure of the social loss. That it is indeed the measure of the social loss can be seen directly, without reference to the loss of consumer's surplus, to the gain in (monopoly or tax) revenue, or to the difference between these, by simply noting that

[5] I am indebted to my colleague Avinash Dixit for discussions that lead to the clarification of this relationship. I am also indebted to Vladimir Simonek of the Prague School of Economics for reminding me that the neglect of the elasticity of supply would be even more troublesome in socialist societies where there is a very strong tendency to neglect the demand side rather than the supply side. Their readers would be disturbed by a diagram like Figure 3 where, at least initially, the falling demand curves are much more prominent than the rising supply curve. It might be useful, therefore, to show the tax revenue curve separately from the demand curve and more clearly as one that depends on *both* the supply and demand curves. This is done in Figure 4 where T, the tax revenue curve, is made up of the vertical distances, taken from Figure 3, between the rising marginal cost or supply curve MC' that passes through E and the falling demand curve P that passes through D. The tax per unit is ED or SR and the tax elasticity is seen directly as BD/DA.

[6] As is claimed, e.g. by Diamond and Mirrlees, p. 14, in terms of an alleged basic asymmetry between demand and supply.

the shaded area represents the social damage from the shift of resources from "here," where the value of their marginal product is between OP_0 and OP_1, to "there" where their marginal product is equal to the (constant) marginal cost OC.

3. *Inputs as Negative Outputs*

The mathematically elegant device of treating inputs as negative outputs obscures the damage caused by the shifting of resources when the tax discourages the producer rather than the consumer. The mathematician may retort that his algebra has included all these effects in his "negative outputs," but rising supply curves are not made flat by calling inputs negative outputs. That eliminates only such increasing MC as is due to diminishing utility of negative inputs (e.g. of "not working"), but it leaves such increasing MC as is due to diminishing transformability of one output into another on account of heterogeneity of productive resources.[7] The treatment of inputs as negative outputs can easily mislead one into taking the asymmetry between subjective preferences and objective production possibilities as an asymmetry between demand and supply, and thereby seduce even economists who are not allergic to consumer's surplus into disregarding such social damage as consists not of the reduction of (consumer-producer's) surplus from negative outputs but of the reduction of *rents* on intra-marginal *nonhuman* productive resources.

The lumping together of *rents* with producer's surplus from *human* inputs, and then transposing both together into consumer's surplus from negative outputs, does indeed translate all possible social damage into the destruction of consumer's surplus. But such a procedure obscures rather than brings out the asymmetry between demand and supply that is fundamentally the difference between subjective (human) preferences (between different outputs or inputs) and objective production feasibilities.

Nevertheless, a careful use of the device of treating inputs as negative outputs can serve to clarify a distinction that is perhaps more fundamental than that between consumer's surplus and producer's surplus. This is the distinction between (a) the social damage from taxation in inducing *subjectively inferior consumption* (including the provision of such inputs, treated as negative outputs, as are regarded as inferior to, i.e. more objectionable than, alternative inputs) and (b) the social damage from taxation in inducing *objectively inferior production* (the use of inputs, between which there are no human preferences, to produce a less valuable output).

The first kind of social damage includes both consumer's surplus and producer's surplus. It consists of that part of the social damage which is caused by the taxation inducing a movement against *human* resistance of consumer-producer's diminishing marginal *subjective* substitutability, which is represented by the convexity (downward) of indifference curves or surfaces, and which may be called *subjective social damage.* The second kind of social damage consists of the social damage which is caused by the taxation inducing a movement against *technical* resistance, diminishing marginal *objective* substitutability, which is represented by the convexity (upward) of the opportunity curves or surfaces, and which may be called *objective social damage.*

In Figure 5 the imposition of a tax on X (but not on Y) causes the equilibrium to move along the production opportunity curve AB (which shows what combinations of X and Y are available for consumers after any using up of resources by the government) from P on indifference curve I to Q on the lower indifference curve I_t. The difference between the slopes of the two

[7] It eliminates them from the MC curve, but they are only translated (in "space" as well as in language) into negative slope in the demand curves for the negative output. The negative outputs would have to be retranslated into positive inputs and the tax on them, based on the inverse of their elasticity of demand retranslated into a tax on the input based on the inverse of their elasticity of supply (unless the Internal Revenue could be educated into imposing the appropriate taxation on the negative outputs in accord with their elasticity of demand). But that would still not cover that part of the increasing MC which is due to technical diminishing transformability of one output into another.

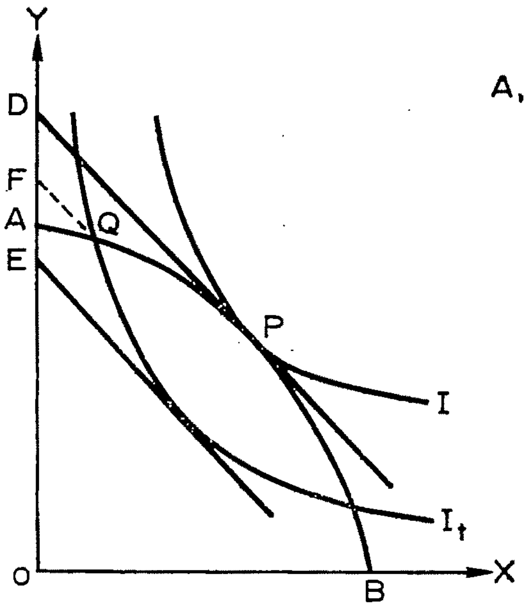


FIGURE 5

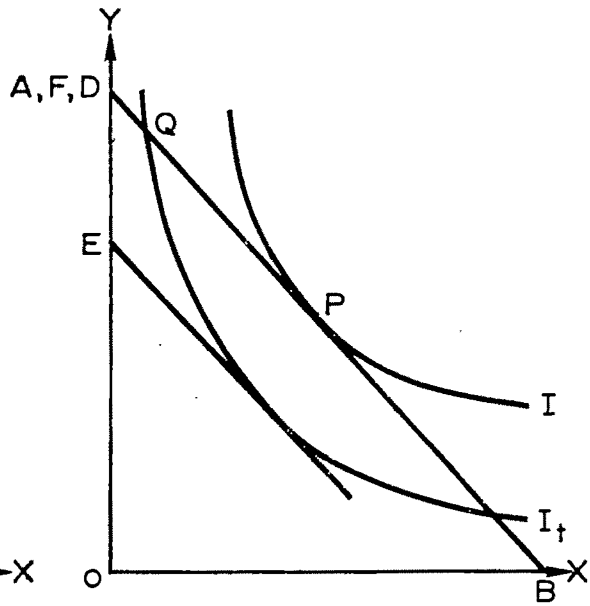


FIGURE 6

curves at Q is equal to the tax rate on X. The social damage is equivalent to an income reduction DE, but the value of the output, which is the income of the owners of the factors of production, will have fallen only by DF (both being measured in terms of Y at the no-tax relative prices).

The amount of social damage is determined in part by the convexity of the indifference curve (or rather the degree to which the slope of successive indifference curves is greater as one moves up the opportunity curve from P to Q) and in part to the convexity of the opportunity curve itself as one moves up it, since both of these affect the position of Q and the level of I_t. The convexity of the opportunity curve is responsible for the reduction DF in income to the owners of the factors of production. This is the *objective* part of the social damage. The convexity of the indifference curves (or the degree to which their slope becomes steeper as one moves up along BA) is responsible for the rest, the *subjective* part of the social damage, FE.

The point may be highlighted by considering the extreme cases. If there are no objective resistances to switching between the production of X and the production of Y, so that the opportunity curve is a straight line, we see in Figure 6 that there is no objective social damage. DF disappears. In this case, if we call inputs involving human preferences negative outputs, we are justi-

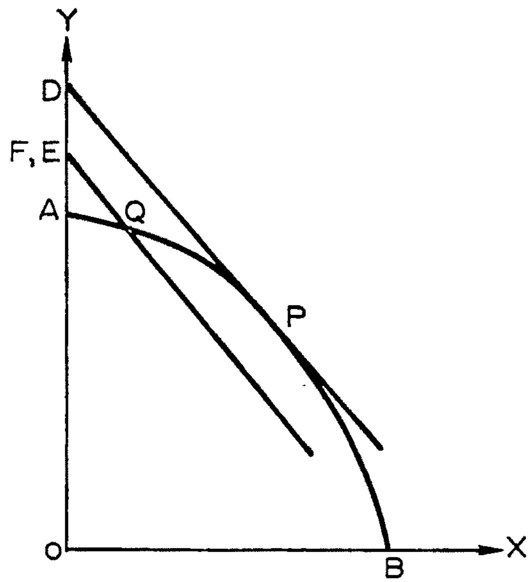


FIGURE 7

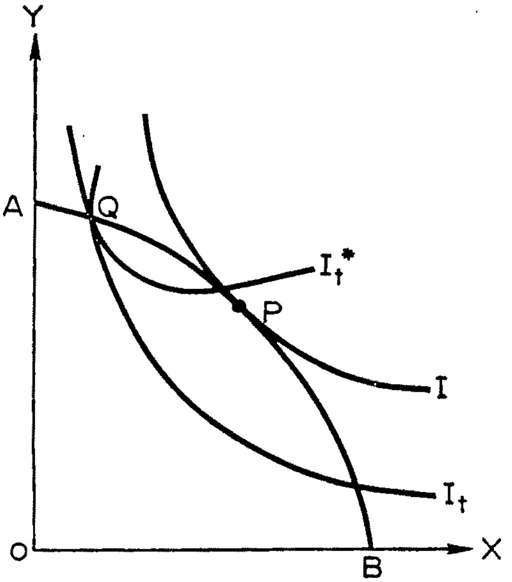


FIGURE 8

fied in assuming constant costs and in either calling all the social damage a loss of consumer's surplus or in saying that the producer's surplus is taken care of together with the consumer surplus.

If there are no subjective resistances to switching between the consumption of X and Y, so that the indifference curves are straight lines, we see in Figure 7 that there is no subjective social damage. FE disappears. The social damage (DE) consists entirely of the decrease (DF) in the income of the owners of the factors of production. The consumers (including the consumers of negative output) are faced by no price changes. They had no consumer's surplus in the first place and have lost none. It is only as owners of the factors of production that they have lost income as the tax induced them to shift against *objective* resistance and their *rents* have diminished by the total amount of the social damage. In both of these extreme cases there is no change in relative prices and thus no "index number" ambiguity about the amount of social damage.

If there are neither subjective nor objective resistances, both the opportunity curve and the indifference curve are straight lines. There is then no social damage at all. If the indifference curves have the same slope as the opportunity curve, a tax on X reduces its output to zero and no tax can be collected. Consumers are not forced onto a lower indifference curve and the owners of factors of production have no reduction in income. If the indifference curves do not have the same slope as the opportunity curve, only one of the two goods is produced, a tax on it can be considered either as a "lump sum" tax or as a "universal" proportional tax and again there is no social damage.[8]

REFERENCES

W. J. Baumol and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *60*, 265-83.

P. A. Diamond and J. A. Mirrlee, "Optimal Taxation and Public Production," mimeo. working paper 22, M.I.T., 1968.

A. P. Lerner, "A Note on Socialist Economics," *Rev. Econ. Stud.*, Oct. 1936, *4*, 72-76.

———, *The Economics of Control.* New York 1944.

[8] The transfer of income from the owners of relatively cheapened productive resources to the owners of those which have gone up in price, and from the consumers of goods that have become more expensive to the consumers of goods that have become cheaper, may so change "collective preferences" as to make P inferior to Q. But this could happen only if the transfer of income brought about a strong shift in preference for Q as compared with P such as would be indicated by the new indifference curve I_t^* in Figure 8 instead of I_t. Since this possibility is no more likely than the opposite, which would *increase* the inferiority of Q to P, it may be disregarded until some positive evidence for it should appear.

On the Optimum Structure of Commodity Taxes

By AVINASH K. DIXIT*

This note is intended to provide a concise algebraic treatment of the problem of optimum commodity taxation, taking explicit account of the possibility that some commodities may be untaxable. This possibility has been mentioned constantly in the literature, and has been analysed at length in the recent paper by Abba Lerner. I shall find support for many of his arguments here, and obtain some precise results on qualitative points made by him.

The analytic framework I shall use is that introduced by Paul Samuelson and elaborated by Peter Diamond and James Mirrlees. The analyses, especially the latter, often become too competent to be comprehensible and I shall simplify the framework to some extent. The basic simplification is to suppose that the government wishes to raise a fixed amount of revenue rather than to command a fixed bundle of commodities. I shall also assume that there is only one consumer in the economy; generalizations based either on Pareto-efficiency considerations or on an individualistic social welfare function add little to the problem of the optimum tax structure.

For the geometrically minded, I shall provide an interpretation of the results in terms of consumers' and producers' surplus. This is, of course, not meant to be taken rigorously. I shall also derive the standard tax formulae, e.g., the proportional reduction formula (Frank Ramsey) and the inverse elasticity formula of Ramsey, A. C. Pigou, and Marcel Boiteux, under special assumptions, and I hope that this will prove to be a step towards clarifying some prevailing confusion on the topic. Doubtless this will *not* prove to be the last word on the subject.

I shall begin by considering just one consumer who faces given supply prices and maximizes utility subject to a budget constraint. As usual, factors supplied by the consumer will be treated as negative demands.[1] It is then possible to allow for many types of labor which the consumer can supply, each with its own wage and disutility. Suppose there are m taxable commodities x_i with prices p_i without tax, the tax per unit being t_i so that the prices paid by the consumer are $P_i = p_i + t_i$. There are n untaxable commodities y_j with prices q_j. We have already taken into account the income the consumer earns from his supply of factors. Suppose the transfer payments to him, if any, amount to I; then his budget constraint is

* At the time the article was written, the author was assistant professor of economics, University of California, Berkeley. He is now a Fellow of Balliol College, Oxford. He would like to take this opportunity to thank Professor Lerner for many helpful discussions on the subject.

[1] This is often made out to be a mathematical trick, but that is far from being the case. If we insist on keeping all quantities positive and apply the term 'demand' both to commodity demands and factor demands (and similarly for 'supply'), we conceal the fact that the former comes from consumers and the latter from producers. With our convention, the term 'demand' applies to all quantities decided by consumers (commodity demands and factor supplies) and 'supply' to all those decided by producers (commodity supplies and factor demands). This enables us to separate the two sources of inefficiency, viz., 'wrong' decisions of consumers and those of producers. If the conventional reader is confused by this usage, he can use the terms consumers' prices and producers' prices instead of demand and supply prices in what follows. I do agree with Lerner, however, that misuse of this convention is possible.

$$\sum_{i=1}^{m} P_i x_i + \sum_{j=1}^{n} q_j y_j = I \tag{1}$$

subject to which he maximises a utility function

$$U(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$$

This yields the first-order conditions for a maximum

$$\partial U/\partial x_i = \lambda P_i \qquad \partial U/\partial y_j = \lambda q_j \tag{2}$$

and the demand functions

$$\begin{aligned} x_i &= x_i(P, q, I) \\ y_j &= y_j(P, q, I) \end{aligned} \tag{3}$$

where, of course, P, q are the vectors (P_i), (q_j), respectively. We can then define the 'indirect utility function'

$$V(P, q, I) = U(x(P, q, I), y(P, q, I)) \tag{4}$$

and it is easy to show that[2]

$$\partial V/\partial P_i = -\lambda x_i \qquad \partial V/\partial q_j = -\lambda y_j \tag{5}$$

where λ is the same as in (2), and can be interpreted as the marginal utility of money.

Now suppose the government wants to raise a fixed amount of revenue, C, by taxing the commodities x_i. This implies

$$\sum_{i=1}^{m} t_i x_i = C \tag{6}$$

To minimize the harm to the consumer, the government must choose t_i subject to (6) in such a way as to maximize V. The first-order conditions for that are

$$\partial V/\partial P_k = \mu\left\{x_k + \sum_{i=1}^{m} t_i \partial x_i/\partial P_k\right\} = -\lambda x_k$$

which simplify to

$$\sum_{i=1}^{m} t_i \partial x_i/\partial P_k = -\nu x_k \tag{7}$$

where $\nu = 1-\lambda/\mu$. The interpretation of ν will be given shortly.

I shall first show how various special cases can be derived from the basic formula (7). If all commodities are taxable, we get by differentiating the budget constraint (1) with respect to P_k,

$$\sum_{i=1}^{m} P_i \partial x_i/\partial P_k + x_k = 0$$

Together with (7), this yields

$$\sum_{i=1}^{m} (t_i - \nu P_i)\partial x_i/\partial P_k = 0 \tag{8}$$

Thus a proportional tax structure ($t_i = \nu P_i$ or $t_i = \nu p_i/(1-\nu)$) is optimum. Moreover, if the matrix $(\partial x_k/\partial P_i)$ is non-singular, it is the only optimum structure.

With all commodities taxable, let $t_i = \alpha p_i$ be the optimum proportional tax structure. Then t_i, x_i and α are all functions of C, and we find

$$\begin{aligned} dx_k/dC &= \sum_{i=1}^{m} (\partial x_k/\partial P_i)(dt_i/dC) \\ &= \frac{1}{\alpha}\frac{d\alpha}{dC}\sum_{i=1}^{m} t_i \partial x_k/\partial P_i \end{aligned}$$

Using the symmetry of the Slutsky matrix, this becomes

$$\begin{aligned} dx_k/dC &= \frac{1}{\alpha}\frac{d\alpha}{dC}\sum_{i=1}^{m} t_i \\ &\quad \cdot\{\partial x_i/\partial P_k - x_i \partial x_k/\partial I + x_k \partial x_i/\partial I\} \\ &= \frac{1}{\alpha}\frac{d\alpha}{dC}\{-\nu x_k - C\partial x_k/\partial I\} \end{aligned}$$

by (6) and (7).

[2] Differentiating (4), we have

$$\begin{aligned} \partial V/\partial P_i &= \sum_{k=1}^{m} (\partial U/\partial x_k)(\partial x_k/\partial P_i) \\ &\quad + \sum_{j=1}^{n} (\partial U/\partial y_j)(\partial y_j/\partial P_i) \\ &= \lambda\left\{\sum_{k=1}^{m} P_k \partial x_k/\partial P_i + \sum_{j=1}^{n} q_j \partial y_j/\partial P_i\right\} \text{ from (2)} \\ &= \lambda\left\{\frac{d}{dP_i}\left[\sum_{k=1}^{m} P_k x_k + \sum_{j=1}^{n} q_j y_j\right] - x_i\right\} \\ &= -\lambda x_i \text{ from (1).} \end{aligned}$$

Thus, when $C=0$, we get

$$\frac{1}{x_k}\frac{dx_k}{dC} = -\frac{\nu}{\alpha}\frac{d\alpha}{dC}, \qquad \text{for all } k \tag{9}$$

This is the Ramsey formula of proportional reduction. It is an approximate result in the neighborhood of zero tax revenue, but does *not* need assumptions regarding independence of demands for its validity.

If the demands are independent, i.e., if $\partial x_k/\partial P_i=0$ whenever $k\neq i$, (7) becomes

$$t_i dx_i/dP_i = -\nu x_i,$$

which may be written

$$t_i/P_i = \nu/\epsilon_i \tag{10}$$

where ϵ_i is the numerical value of the elasticity of demand for the *i*th commodity. This is the inverse elasticity formula. In the current context, however, this takes a very peculiar form. For, if all commodities are taxable, it is easy to see by differentiating (1) that independent demands imply $\epsilon_i=1$ for all i. The optimum structure is proportional, after all. In fact, (8) tells us that it must be so, since the matrix $(\partial x_k/\partial P_i)$ is now diagonal.[3]

Let us turn to the case where there are untaxable commodities. Differentiating (1) in this case, we get

$$x_k + \sum_{i=1}^{m} P_i \partial x_i/\partial P_k + \sum_{j=1}^{n} q_j \partial y_j/\partial P_k = 0$$

and then

$$\sum_{i=1}^{m} (t_i - \nu P_i)\partial x_i/\partial P_k = \nu \sum_{j=1}^{n} q_j \partial y_j/\partial P_k$$

If the matrix $(\partial x_i/\partial P_k)$ is nonsingular,[4] and its inverse is written (A_{ik}), we have the formula

$$t_i = \nu\left\{P_i + \sum_{j=1}^{n}\sum_{k=1}^{m} q_j(\partial y_j/\partial P_k)A_{ik}\right\} \tag{11}$$

This structure is, needless to say, not proportional in general. A sufficient condition for proportionality is that $\partial y_j/\partial P_k=0$ for all j and k. This says that no income is diverted into the untaxed commodities as a result of a change in any of the taxes. This does not require a perfectly inelastic demand for any y_j (corresponding to a perfectly inelastic supply of labor when leisure is untaxable).

If, with some commodities untaxable, we have independent demand for the taxed commodities, i.e., $\partial x_i/\partial P_k=0$ whenever $k\neq i$, we get back the inverse elasticity formula, which is now meaningful since the elasticities can be arbitrary. Such a structure of demand does require all the burden of adjustment to fall on the untaxed commodities when the amount spent on x_i changes in response to a change in P_i. This is surely just as unreasonable a structure of demand as the one for which no y responds to a change in any of the P's. In the general case, (11) shows that the optimum structure is a mixture of proportional and nonproportional taxes, although I have not been able to obtain a simple weighted average formula along the lines suggested by Lerner.

If demands are independent and linear, we have

$$t_i a_i = -\nu x_i$$

where $a_i = dx_i/dP_i$. The reduction in the amount of x_i is

$$\begin{aligned}\Delta x_i &= -a_i \Delta P_i\\ &= -a_i t_i\\ &= \nu x_i\end{aligned}$$

In this case, the proportional reduction formula is exact. (If all commodities are taxable, independent demands cannot be linear since they must have (unitary) constant elasticities.)

[3] Unless some $\partial x_i/\partial P_i$ is zero (completely inelastic demand). In this case, we shall get a corner solution for the tax structure This result is too well-known to need further comment.

[4] This is one case where it is better not to express the formulae using the Slutsky matrix of compensated demand derivatives since that matrix is always singular.

I shall now interpret the basic formula (7) in terms of the change in consumers' surplus. Suppose we have an optimum structure, and consider a small deviation from it by changing the ith tax rate t_i, and with p_i fixed, the ith demand price P_i, by an amount P_i. This causes a change in the amounts bought of all goods, with

$$\Delta x_k = (\partial x_k/\partial P_i)\Delta P_i$$

For the ith good, the change in consumers' surplus is given by $x_i\Delta P_i$. In the virtual displacement we are considering, the price to the consumer of all other goods is unchanged. Since each amount x_k changes, however, there is a change in the tax revenue. The total change in the revenue is given by

$$x_i\Delta t_i + \sum_{k=1}^{m} t_k\Delta x_k = \left\{x_i + \sum_{k=1}^{m} t_k\partial x_k/\partial P_i\right\}\Delta P_i$$

Figure 1 illustrates all this. The excess of the change in the consumers' surplus over the change in the tax revenue is the 'deadweight loss' or the 'marginal social damage.'

The change in surplus per unit of extra tax revenue is then the ratio

$$x_i \bigg/ \left\{x_i + \sum_{k=1}^{m} t_k\partial x_k/\partial P_i\right\}$$

From (7), we see that this ratio is required to be $1/(1-\nu)$ at the optimum. Thus the condition requires the change in any tax rate to yield the same change in consumer's surplus per unit of additional tax revenue, and we see that this common value is given by $1/(1-\nu)$, thus giving an interpretation to ν. It is the ratio of the marginal deadweight loss to the marginal loss in consumers' surplus. Thus it can serve as a measure of the inefficiency of the tax. A better measure is $\nu/(1-\nu)$, the marginal loss/revenue ratio.

This condition is very much like the equalization of the ratios of average and marginal tax rates proposed by Lerner, the only difference being that he considers small deviations in the quantity x_i as the starting point of the argument. The two yield equivalent results in most cases, but starting with the price is perhaps slightly preferable from the point of view of the theory.

It is worth mentioning that in all this

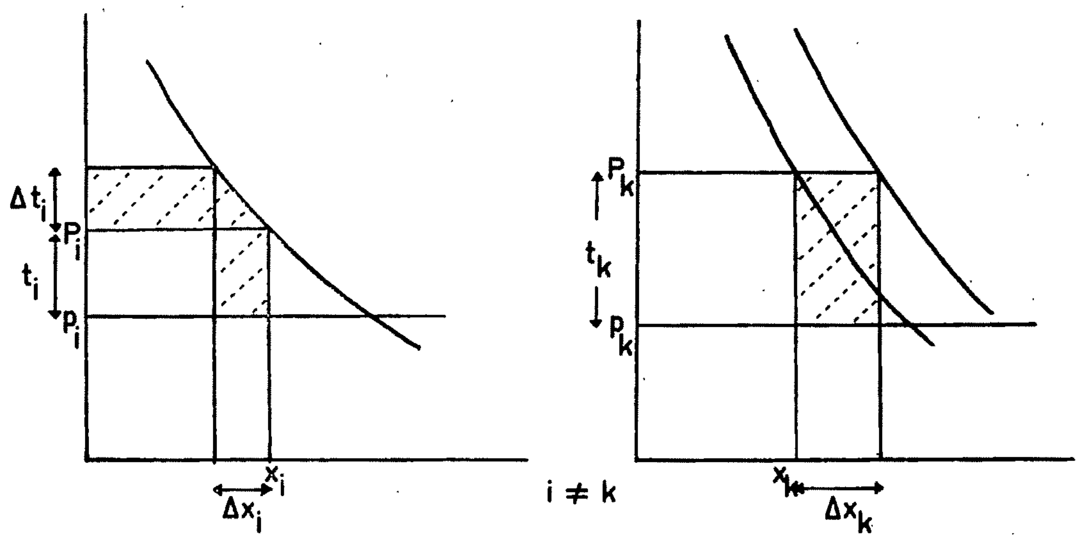


FIGURE 1

argument we have no need to consider the change resulting from the change in the quantities of the untaxable goods. Since their demand and supply prices are equal in the optimum structure, the dead-weight loss from them is a second-order quantity.

This analysis proceeds on the assumption of fixed supply prices, and thus neglects the possibility that a tax structure may distort efficiency by causing a wrong allocation of resources in production. This can be remedied by introducing a production function in the model and determining supply prices by profit-maximization; then they can vary with output structure. I shall confine myself to the case where only one good is not taxable. This involves no essential loss, since we can always normalize the prices to have zero tax on one good when there is production, and the case where more than one good is untaxable follows the same lines to the same results with some additional algebra. The transformation curve solved for this good, y, then becomes the production function

$$y = g(x_1, x_2, \ldots, x_m) \tag{12}$$

If y is chosen as numeraire, the profit is given by

$$y + \sum_{i=1}^{m} p_i x_i$$

and its maximization gives supply prices as functions of x_i:

$$p_i = -\frac{\partial g}{\partial x_i}(x_1, x_2, \ldots, x_m) \tag{13}$$

The reader can verify that with our sign convention for inputs, every $\partial g/\partial x_i$ is negative, so every p_i is positive. If y is an output and x_i an input, p_i has the interpretation of the marginal product; if x_i is an output and y an input, p_i is the marginal cost. If both are outputs (inputs), it is the marginal rate of transformation (substitution).

We must now maximize $V(P, 1, I)$ subject to the constraint

$$\sum_{i=1}^{m} (P_i - p_i)\, x_i = C \tag{14}$$

and taking into account all the functional relationships among the variables. Since (13) gives p_i as functions of the x_i, and (3) gives x_i as functions of the P_i, it is convenient to take P_i as the independent variables. The relationships, however, are fully circular in the standard equilibrium manner and any starting point will do just as well in most cases (some singularities can cause minor difficulties). The best causal interpretation is one where the tax rates t_i are announced first, and P_i, p_i and x_i all adjust until the difference $P_i - p_i$ equals t_i. The latter are, of course, calculated in advance so as to satisfy the revenue constraint.

The first-order conditions for maximization are

$$0 = \partial V/\partial P_k = \mu\left\{x_k + \sum_{i=1}^{m} (P_i - p_i)\partial x_i/\partial P_k - \sum_{i=1}^{m}\sum_{h=1}^{m} x_i(\partial p_i/\partial x_h)(\partial x_h/\partial P_k)\right\}$$

Evaluation of $\partial V/\partial P_k$ is slightly more difficult since (5) does not apply. In equilibrium, y is related to (x_i) via (12), and this has to be taken into account. In fact we have

$$\begin{aligned}
\partial V/\partial P_k &= \sum_{i=1}^{m} (\partial U/\partial x_i)(\partial x_i/\partial P_k) + (\partial U/\partial y)(\partial y/\partial P_k) \\
&= \lambda \sum_{i=1}^{m} P_i \partial x_i/\partial P_k + \lambda \sum_{i=1}^{m} (\partial g/\partial x_i)(\partial x_i/\partial P_k) \\
&= \lambda \sum_{i=1}^{m} (P_i - p_i)\partial x_i/\partial P_k \text{ using (13).}
\end{aligned}$$

This enables us to write the condition as

(15)
$$-\nu x_k = \sum_{i=1}^{m}\left\{t_i - \nu \sum_{h=1}^{m} x_h \partial p_h/\partial x_i\right\} \partial x_i/\partial P_k$$

Note that to do this, we must interchange the dummy indices i and h in the double sum. Finally, substituting from the derivative of the budget constraint with respect to P_k, we get

$$t_i = \nu\left\{P_i + \sum_{k=1}^{m} \partial y/\partial P_k A_{ki} + \sum_{h=1}^{m} x_h \partial p_h/\partial x_i\right\} \tag{16}$$

This formula differs, apart from the specialization to the case of one untaxable good, from (11) in that it involves derivatives of supply functions on the right-hand side. Thus, under general conditions (subject only to an interior maximum) production can affect the optimum tax structure.

This dependence vanishes, however, if we assume constant returns to scale in production. (Note that this is *not* the same as assuming infinitely elastic supply curves: constant returns to scale are perfectly compatible with diminishing returns to one factor, and in that case (13) shows that the supply curve depicting p_i against x_i has a positive slope.) If g is homogeneous of degree one in the x_i, then every $\partial g/\partial x_i$ is homogeneous of degree zero in them, and

$$\begin{aligned}\sum_{h=1}^{m} x_h \partial p_h/\partial x_i &= -\sum_{h=1}^{m} x_h(\partial^2 g/\partial x_h \partial x_i)\\ &= -\sum_{h=1}^{m} x_h(\partial^2 g/\partial x_i \partial x_h)\\ &= \sum_{h=1}^{m} x_h(\partial p_i/\partial x_h)\\ &= 0 \quad \text{by Euler's theorem.}\end{aligned}$$

In this case, (16) reduces to (11). This special case of our analysis confirms the result obtained by Diamond and Mirrlees.[5]

The formula for the optimum tax structure with production, too, can be given an interpretation similar to that given before. If we contemplate a change of ΔP_i in P_i, the accompanying changes in other variables will be as follows:

$$\Delta x_k = (\partial x_k/\partial P_i)\Delta P_i$$

$$\Delta p_k = \sum_{h=1}^{m} (\partial p_k/\partial x_h)(\partial x_h/\partial P_i)\Delta P_i$$

$$\Delta t_k = \begin{cases}\Delta P_i - \Delta p_i & k = i\\ \quad -\Delta p_k & k \neq i\end{cases}$$

The change in tax revenue is $\sum_{k=1}^{m} t_k \Delta x_k + \sum_{k=1}^{m} x_k \Delta t_k$, which, written in full, becomes

$$\left\{x_i + \sum_{k=1}^{m} t_k \partial x_k/\partial P_i - \sum_{k=1}^{m}\sum_{h=1}^{m} x_k(\partial p_k/\partial x_h)(\partial x_h/\partial P_i)\right\}\Delta P_i$$

The change in the consumers' and producers' surplus is simply

$$\sum_{k=1}^{m} x_k(\Delta P_k - \Delta p_k) = \sum_{k=1}^{m} x_k \Delta t_k$$

and it is easy to see that the requirement that the change in the surplus per unit of change in revenue be the same for changes in any P_i, yields the same conditions as (15).

If demands *and* supplies are independent, (15) becomes

$$t_k = \nu\{x_k dp_k/dx_k - x_k dP_k/dx_k\}$$

Dividing by P_k and converting the first term into an elasticity,

$$t_k/P_k = \nu\left\{(p_k/P_k)\left(\frac{x_k}{p_k}\frac{dp_k}{dx_k}\right) - \frac{x_k}{p_k}\frac{dP_k}{dx_k}\right\}$$

[5] I have not been able to find a simple intuitive reason why constant returns to scale lead to this. It seems that the total change in producers' surplus caused by the shifting of all supply curves has to be zero in that case. Lerner appears to have overlooked this interdependence by looking only at the partial picture of supply for one commodity.

or, writing τ for the tax rate as a percentage of the demand price, ϵ_s for the supply elasticity and ϵ_d for the demand elasticity

$$\tau = \nu\{(1-\tau)/\epsilon_s + 1/\epsilon_d\}$$

which becomes, on solving for τ,

$$\tau = \frac{(\epsilon_s + \epsilon_d)}{\epsilon_d(\nu + \epsilon_s)} \tag{17}$$

This replaces (10) when there is production. This gives a precise formulation of the idea of a "*tax rate* proportional to a *combination* of the elasticity of supply and the elasticity of demand" mentioned by Lerner (p. 290). The two do not get equal weights since they are not evaluated at the same price. Moreover, the difference in the two prices involves the tax, so the final formula is not a simple arithmetic combination of the inverses of the elasticities.[6]

The sharp reader must surely have noticed that although there are profits in equilibrium with nonconstant returns to scale, we have not taken them into account. Possibility of profits, however, brings with it the possibility of a profit tax and an accommodation of it in the optimum structure. That must remain outside the scope of this short note.

[6] Pigou mentions taxes proportional to $1/\epsilon_d + 1/\epsilon_s$, refering to Ramsey for proof. Ramsey derives the correct formula, which for small taxes reduces to Pigou's as an approximation.

REFERENCES

M. Boiteux, "Sur la gestion des Monopoles Public astreints à l'équilibre budgétaire," *Econometrica,* Jan. 1956, *24,* 22-40.

P. A. Diamond and J. M. Mirrlees, "Optimal Taxation and Public Production," mimeo. working paper No. 22, M.I.T., May 1968.

A. P. Lerner, "On Optimal Taxes with an Untaxable Sector," *Amer. Econ. Rev.,* June 1970, *60,* 284-94.

A. C. Pigou, *A Study in Public Finance,* 3rd ed., London 1947.

F. P. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.,* Mar. 1927, *37,* 47-61.

P. A. Samuelson, "Theory of Optimal Taxation," unpublished 1951.

A Diagrammatic Exposition of Optimal Growth

By NISSAN LIVIATAN*

The theory of optimal growth can be analyzed, as is well known, by using optimal-control methods or by applying the technique of dynamic programming. It seems, however, that the graphical expositions for simple growth models have been developed only for the former approach. Thus it is customary to present the optimal growth path in terms of a trajectory in the phase-space.[1] However, as far as I know, there has been no attempt to formulate a graphical analysis which will bring out the essence of the dynamic programming approach[2] to optimal growth.

The purpose of this paper is to fill this methodological gap. Moreover, it will be seen that the graphical presentation of the dynamic programming approach is closely related to the traditional Fisherine diagram of intertemporal analysis and to the standard Hicksian tools of demand analysis. Indeed, we intend to show that the exposition of the optimal time path in a simple growth model does not require much more than a dynamic version of the well-known two-period Fisherine diagram, with current consumption on one axis and next year's capital on the other.

I. *A Restatement of Some Elementary Results*

In our analysis, which is based on a discrete time model, we shall deal with a one-sector model with no technological progress. The production function of the economy is assumed to be of the following form:

$$(1)\qquad S_{t+1} + C_{t+1} = F(S_t, L_t)$$

where S_t denotes the stock of capital which enters as an input in the production function in period t, L_t is labor input in period t, and C_{t+1} is consumption in period t+1. (We may assume that (1) incorporates an exponential depreciation factor.) An alternative way of writing (1), which is more useful for our purposes, is as follows. Define K_t as the capital stock available (for production or consumption) in the beginning of period t. We then have

$$(2)\qquad K_t = S_t + C_t$$

Substituting (2) into (1) we obtain

$$(3)\qquad K_{t+1} = F(K_t - C_t, L_t)$$

It should be noted that $F_1 = \partial F/\partial(K_t - C_t)$ is the marginal product of capital in period t in producing the capital stock of period t+1. Hence $(F_1 - 1)$ is the net own rate of return on capital, which we shall assume to be nonnegative.

We shall assume as usual that F is subject to constant returns to scale with respect to the inputs $(K_t - C_t)$ and L_t and that L_t grows according to

$$(4)\qquad L_{t+1} = nL_t \qquad n > 1, \quad \text{for all t}$$

* The author is professor of economics at the Hebrew University of Jerusalem and visiting professor at University of California, Berkeley. This study was done during the tenure of a Ford Foundation Fellowship.

[1] See, for example, the presentation in David Cass (pp. 236–38) which is based on control theory, or the graphical exposition by Tjalling C. Koopmans based on an equivalent calculus-of-variations formulation.

[2] For an analytic exposition of optimal gowth based on dynamic programming see, for example, Roy Radner.

where n is given exogenously. Using the property of constant returns to scale, we may divide all variables in (3) by L_t, which yields

$$\frac{K_{t+1}}{L_t} = F\left(\frac{K_t}{L_t} - \frac{C_t}{L_t},\ 1\right). \tag{5}$$

Dividing both sides of (5) by n and using (4), we have

$$\frac{K_{t+1}}{L_{t+1}} = \frac{1}{n} F\left(\frac{K_t}{L_t} - \frac{C_t}{L_t},\ 1\right). \tag{6}$$

Using lower case letters to denote per capita variables, we may write (6) as

$$k_{t+1} = \frac{1}{n} F(k_t - c_t,\ 1). \tag{7}$$

As is usual in growth theory, we make the following assumptions about $F(k_t - c_t, 1)$:

$$F(0, 1) = 0 \tag{8}$$

$$F_1 \geq 1, \qquad F_{11} < 0 \quad \text{for all } k_t - c_t > 0 \tag{9}$$

$$F_1(0, 1) = \infty, \qquad F_1(\infty, 1) = 1 \tag{10}$$

where F_{11} denotes $\partial^2 F/\partial(K_t - C_t)^2$. Since n is a constant throughout our analysis, it is convenient to define a new function

$$f(k_t - c_t) \equiv \frac{1}{n} F(k_t - c_t,\ 1) \tag{11}$$

where f incorporates implicitly the growth factor n. Applying assumptions (8)–(10) to f, we may write

$$f(0) = 0; \tag{8'}$$

$$f' \geq \frac{1}{n}, \qquad f'' < 0; \tag{9'}$$

$$f'(0) = \infty, \qquad f'(\infty) = \frac{1}{n} \tag{10'}$$

where f' and f'' denote the first- and second-order derivatives of f. Note that in assumptions (8′)–(10′) we made use of relation $f' = F_1/n$. (It may be pointed out

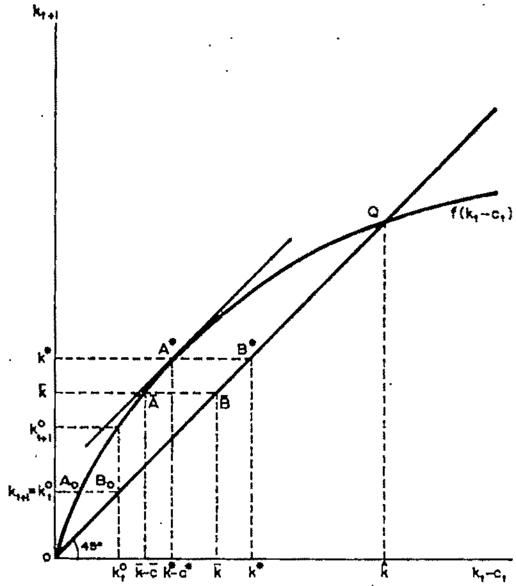


FIGURE 1

that our analysis will remain unaltered if we use the more general assumption $0 \leq f'(\infty) < 1$. This will include as a special case the assumption $f' \geq 0$ and $f'(\infty) = 0$, implying a negative *net* marginal product of capital for a sufficiently large capital input.)

The function $f(k_t - c_t)$ is illustrated in Figure 1. It should be noted that by assumption (10′) the curve corresponding to f must be above the 45° line for some range of $(k_t - c_t)$ and that it must intersect the latter at some $k_t - c_t > 0$, since $f'(\infty) = 1/n < 1$. It follows from the foregoing facts that for any value of k_t, $0 < k_t < \hat{k}$, there corresponds a positive value of c_t which will leave capital intact.

To see this, consider $k^0_t - c_t$ in Figure 1. If $c_t = 0$ then $k^0_t - c_t = k^0_t$ and we have $k_{t+1} = k^0_{t+1} > k^0_t$ so that capital increases. However, if c_t equals the segment A_0B_0, then $k_{t+1} = k^0_t$ and capital remains stationary. Denote the relationship between k_t and the value of c which leaves it intact by $c = s(k)$. In view of Figure 1, it is clear that c, considered as a function of k, possesses a

unique positive maximum[3] at some k^*, $0<k^*<\hat{k}$. It is seen immediately from Figure 1 that this maximum corresponds to the point A^* where the tangent to f is parallel to the 45° line. This implies that at A^* we have

$$(12)\qquad f'(k-c)=\frac{F_1(k-c,1)}{n}=1$$

or $F_1=n$. In other words, when stationary consumption per capita is maximized (i.e., at $k=k^*$ and $c=c^*$), the net own rate of return on capital (F_1-1) must equal the net rate of population growth $(n-1)$. This is the so-called "golden rule" of capital accumulation, with c^* and k^* being the golden rule[4] values of c and k.

Another implication of Figure 1 is derived from the fact that the average product of capital is less than unity for $k_t>\hat{k}$, which means that $k_{t+1}<k_t$ if $k_t>\hat{k}$. This implies that for any feasible program (with $0\leq c_t\leq k_t$) k_t and c_t are bounded from above by max $[k_o, \hat{k}]$ for all t.

Let us turn now to the utility function which is supposed to represent the preferences of the economy as a whole. We assume the standard type of utility function used in growth theory, namely

$$(13)\qquad U=\sum_{t=0}^{\infty}a^t u(c_t)$$

where a is a constant subjective discount factor satisfying $0<a<1$, and u is a function of *per capita* consumption. It is assumed that u has the following properties:

$$(14)\qquad u'>0,\qquad u''<0\quad \text{for } c_t>0$$

and in addition, to exclude corner solutions for c_t (where $c_t=0$), we shall assume

$$(15)\qquad \lim_{c_t\to 0} u'(c_t)=\infty.$$

The problem of determining the optimal program can then be stated as follows:

$$(16)\qquad \text{Maximize } U=\sum_{t=0}^{\infty}a^t u(c_t) \text{ with respect to } c_0, c_1, \ldots, \text{ subject to } k_{t+1}=f(k_t-c_t),\ 0\leq c_t\leq k_t,\ t=0,1,\ldots, \text{ given initial capital } k_0.$$

The first-order conditions, or the "Fisher conditions," for the optimal path require the equalization of the marginal rates of substitution in production to those in consumption for every t, which yields

$$(17)\qquad \frac{u'(c_t)}{au'(c_{t+1})}=f'(k_t-c_t)\qquad t=0,1,\ldots.$$

This can be verified as follows. Consider an optimal program and denote the optimal values of k_t by $\bar{k}_t$. If, for some t, we take $\bar{k}_t$ and $\bar{k}_{t+2}$ as given, then in the optimal program $a^t u(c_t)+a^{t+1}u(c_{t+1})$ must be maximized with respect to c_t and c_{t+1} subject to $k_{t+1}=f(\bar{k}_t-c_t)$ and $\bar{k}_{t+2}=f(k_{t+1}-c_{t+1})$, or $\bar{k}_{t+2}=f[f(\bar{k}_t-c_t)-c_{t+1}]$. The latter is a transformation curve between c_{t+1} and c_t with a slope of $f'(k_t-c_t)$. At an optimum of the miniature system we must therefore have an equality of $f'(k_t-c_t)$ with the marginal rate of substitution between c_{t+1} and c_t, i.e., with $u'(c_t)/au'(c_{t+1})$. Since this must be true for all t, we have established (17).

Ordinarily these are only the necessary conditions for an interior maximum. However, under our assumptions about the discount factor and about the concavity of the production and utility functions, it follows that the Fisher conditions are sufficient as well, provided the program satisfies the condition $a^t u'(c_t)\to 0$ as $t\to\infty$ (as, for example, in the case where c_t tends

[3] This can also be verified as follows: In stationary solutions we have $k=f(k-c)$. Differentiation with respect to k and c yields $dc/dk=s'(k)=(f'-1)/f'$. (Note that by (10′) we have $s'(0)=1$.) Differentiating once more we obtain $s''(k)=f''/(f')^2<0$ which shows that $s(k)$ is concave. By assumption (10′) $(f'-1)$ changes sign. It follows therefore that $s(k)$ has a unique absolute maximum. At the maximum point we have $s'(k)=0$, i.e., $f'=1$ as in (12).

[4] See E. S. Phelps.

to a positive stationary limit). Moreover, the program satisfying the foregoing conditions is unique. The proof of these statements[5] is of no direct concern to our discussion and we shall therefore take them as given, or treat them as assumptions.

An important feature of the solution of (16) is that it will remain optimal (as far as present and future consumption is concerned) when the economy reexamines it as it moves actually into the future. This "dynamic consistency" property of the optimal program[6] follows directly from the form of the utility function in (13). A related aspect of (16) is that the solution is independent of the calendar time. The index t should be interpreted as indicating the "number of periods ahead" at any given calendar date. Thus the process is a stationary one.

We have formulated the maximization problem subject to a given arbitrary level of initial capital. As a preliminary inquiry we may, however, disregard initial capital and examine whether our system is at all capable of a stationary optimal solution where $\bar{c}_t = c$ and $\bar{k}_t = k$ for all t. Substituting the constant values of c and k into (17) we obtain

$$(18) \qquad 1/a = f'(k - c).$$

We know that there must exist a positive solution to (18) if $f'(0) > 1/a > f'(\infty)$. Since $1/a > 1$ this is guaranteed by our foregoing assumption (10′). It also follows from the fact that f' is monotonically decreasing that $(k-c)$ is *uniquely* determined by (18). In a stationary solution we must also have

$$(19) \qquad k = f(k - c).$$

We may therefore determine from (18) and (19) a unique solution for c and k individually. The foregoing values, to be denoted by $\bar{c}$ and $\bar{k}$, constitute an optimal stationary solution. Using the fact that $1/a = f'(\bar{k} - \bar{c}) > 1$, we may infer that $\bar{k}$ and $\bar{c}$ are smaller than the corresponding golden rule values, i.e., $\bar{k} < k^*$ and $\bar{c} < c^*$. This is illustrated in Figure 1 where the optimal stationary solution is represented by the point $\bar{A}$ which corresponds to $f' > 1$. The optimal stationary consumption level is then $\overline{AB} < A^*B^*$. It is also seen that as time preference decreases, so that "a" approaches unity, the optimal stationary values approach those of the golden rule.[7]

II. *The Reduced Maximization System*

We shall assume that for any k_o there exists a unique solution of (16). As an indication that our assumption about the existence of a maximum is a reasonable one, we note that the utility sum of any feasible consumption sequence is bounded from above by $u(M)/(1-a)$, where $M = \max [\hat{k}, k_o]$. Consider the maximized value of U, say $\bar{U}$ as a function of k_o, and denote this function by $\bar{U} = v(k_0)$.[8] We then assume that $v(k_o)$ is continuous and twice differentiable for all $k_o > 0$.

Consider a two-stage maximization of (16). In the first stage we maximize the utility function treating c_o and k_1 as given parameters, ignoring completely k_o and the relation $k_1 = f(k_o - c_o)$. We may then write (16) as

$$(20) \qquad u(c_o) + a \max_{c_{t+1}} \sum_{t=0}^{\infty} a^t u(c_{t+1})$$

for a given value of k_1.

Note, however, that the problem of maximizing $\sum_{t=0}^{\infty} a^t u(c_{t+1})$ for a given value of k_1 is formally identical with maxi-

[5] These statements can be verified by adopting the method of proof used by Cass (p. 235) to our model. Note that his equation (11) is our condition on $a^t u(c_t)$.

[6] On the possibility of dynamic inconsistency, see the analysis by R. H. Strotz.

[7] The case where strictly $a = 1$ raises some mathematical problem concerning the convergence of the utility sum. It has been shown, however, by Koopmans that this problem can be overcome by a proper adjustment of the origin of the function u.

[8] In the logarithmic model where $u(c) = \log c$, and $\log f(k_t - c_t) = \beta + \gamma \log (k_t - c_t)$, $0 < \gamma < 1$, one can verify that $v(k_o) = (1/1 - \gamma a) \log k_o +$constant (note that in this model $f'(\infty) = 0$). See, for example, Radner (pp. 96–105).

mizing $\sum_{t=0}^{\infty} a^t u(c_t)$ for a given value of k_o. We may then use the foregoing v function to write

$$(21)\quad \max_{c_{t+1}} \sum_{t=0}^{\infty} a^t u(c_{t+1}) = v(k_1).$$

Substituting in (20), we have

$$(22)\quad \max_{c_{t+1}} U = u(c_o) + av(k_1).$$

This completes the first stage of maximization, from which we have obtained the reduced utility function (22). (It is understood, of course, that the reduced utility function depends not only on the original utility function but also on the production function.)

In the second stage we treat c_o and k_1 as endogenous variables and maximize the reduced utility function (22) subject to the present period's production constraint $k_1 = f(k_o - c_o)$. This determines the current period's optimal values of c_o and k_1. The second-stage problem[9] can then be written as

$$(23)\quad \begin{aligned} &\max_{c_o, k_1} [u(c_o) + av(k_1)] \\ &\text{subject to } k_1 = f(k_o - c_o), \text{ and given } k_o. \end{aligned}$$

The recursive nature of this system is clear. Thus the value of k_1 determined by (23) becomes the next period's k_o which is used to determine by means of the same functions the value of the new k_1 and c_o and so on. Note also that by the definition of v, (23) satisfies

$$(24)\quad v(k_o) = \max_{c_o} \{u(c_o) + av[f(k_o - c_o)]\}$$

which is the fundamental functional equation of dynamic programming.[10]

[9] It should be noted that while (23) represents a two-period analysis, it presupposes that an appropriate infinite horizon problem has been solved in the first stage of maximization. This, however, does not cause any difficulty as long as we are not interested in the computational aspect of the problem.

[10] See discussion concerning this type of equation in R. Bellman.

As a necessary condition for a maximum of the right-hand side of (24), we have

$$(25)\quad u'(c_o) = av'[f(k_o - c_o)]f'(k_o - c_o).$$

Denote the maximizing value of c_o by $\bar{c}_o$, where the latter can be considered as a function of k_o. Substituting in (24), we have

$$(26)\quad v(k_o) = u(\bar{c}_o) + av[f(k_o - \bar{c}_o)].$$

Differentiating (26) with respect to k_o and using (25), we obtain

$$(27)\quad \begin{aligned} v'(k_o) &= u'(\bar{c}_o)d\bar{c}_o/dk_o + av'[f(k_o - \bar{c}_o)] \\ &\quad \cdot f'(k_o - \bar{c}_o)(1 - d\bar{c}_o/dk_o) \\ &= u'(\bar{c}_o) \end{aligned}$$

i.e. the marginal utility of current capital equals the marginal utility of current consumption (and consequently $v' > 0$). This is only natural, since one of the alternatives of using an increase in k_o is consuming it in the current period.

It follows from (27) and (14) that v is concave if $d\bar{c}_o/dk_o > 0$. Consider the hypothesis that $d\bar{c}_o/dk_o \leq 0$ for all k_o. Then, if k_o increases, $\bar{c}_o$ will not increase and hence k_1 will increase, implying that next period's "initial" capital increases. When we are actually in the next period, we face again the same kind of situation (i.e., an increase in initial capital) as in the original period. Repeating the previous argument, we shall find that no c_t will increase, compared with the original path, in spite of the fact that an increase in some c_t's is feasible.[11] Clearly this cannot happen with optimal programs. In particular it contradicts our earlier conclusion that $v' > 0$. By using a somewhat more sophisticated argument, one could also reject the hypothesis that $d\bar{c}_o/dk_o \leq 0$ holds for *some* k_o. Hence $d\bar{c}_o/dk_o > 0$ and $v'' < 0$ for all k_o.

Using the foregoing results and the functional equation, we may infer that

[11] Note that we have used here the property of dynamic consistency which enables us to relate the planned path to the actual one.

$d\bar{c}_o/dk_o<1$. Thus differentiating (25) with respect to k_o and rearranging terms, we obtain

$$(28)\quad \frac{d\bar{c}_o}{dk_o}=\frac{1}{1+\dfrac{u''}{a(v'f''+f'^2v'')}}$$

which is between zero and one. It also follows from these results that $d\bar{k}_1/dk_0=f'\cdot(1-d\bar{c}_o/dk_o)>0$.

III. *The Diagrammatic Analysis of Optimal Growth*

The short-run equilibrium at some given date is illustrated in Figure 2 by means of a Fisherine diagram, where on the horizontal axis we have c_o and on the vertical axis k_1. The production possibility frontier for initial capital k_o is given by AB which represents the function $k_1=f(k_o-c_o)$ in the (c_o, k_1) plane. The slope of this curve at any point is given by $\partial f/\partial c_0=-f'<0$. Similarly, $\partial^2 f/\partial c_o^2=f''<0$. The transformation curve relating k_1 and c_o is then of the usual concave shape. The indifference curves II' in Figure 2 correspond to the family

$$u(c_o)+av(k_1)=\text{constant}.$$

From our earlier results we know that $v'>0$ and $v''<0$, just as is the case with the derivatives of $u(c_o)$. It follows therefore that the indifference curves of our reduced utility function have the usual convex shape. If the initial capital is k_o^o in Figure 2, then the short-run equilibrium is determined at E_o so that the equilibrium values are given by (c_o^o, k_1^o).

Suppose now that initial capital increases to k_o^1. Then the production frontier moves uniformly[12] to the right, and is represented by (say) A_1B_1 in Figure 2. What can be said about the new short-run equilibrium point E_1? We know from our foregoing analysis that both $d\bar{c}_o/dk_o$ and $d\bar{k}_1/dk_o$ are positive, i.e., the "Income Consumption Curve" EE' in Figure 2 must be upward sloping.[13] Hence E_1 must be above E_o and to the right of it.

So far we dealt with short-run comparative statics. We have to determine now how the system actually moves from one time period to the next. Suppose that the economy is initially at E_o in Figure 3. Then in the next period the production frontier moves to (say) A_1B_1. Consider now the intersection point Q_o of the horizontal line corresponding to $\bar{k}_1$ with A_1B_1. The value of consumption corresponding to this point, say c', or ON in Figure 3, is determined by $\bar{k}_1=f(k_o^1-c')$ where $k_o^1=\bar{k}_1$ is next period's initial capital. We may then write

$$(29)\qquad \bar{k}_1=f(\bar{k}_1-c')$$

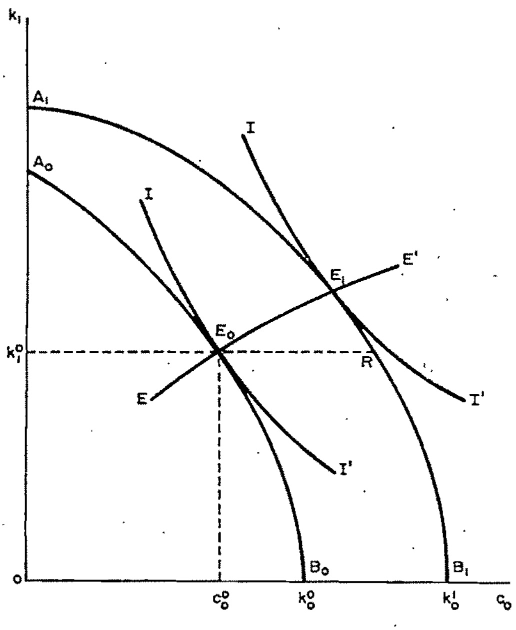


FIGURE 2

[12] Moreover, it can be seen that the production frontier moves to the right parallel (horizontally) to itself. Thus, for example, the slope of the production frontier at R is the same as at E_o. To see this, note that both E_o and R correspond to the same value of k_1, i.e. to k_1^o. Since k_1 is a monotonic function of k_o-c_o, this means that both E_o and R correspond to the same value of k_o-c_o, which implies that $f'(k_o-c_o)$ is also the same.

[13] In the logarithmic model (described in fn. 8) the formula for the income consumption curve is given by $k_1=Ac_o^\gamma$ where A is a positive constant.

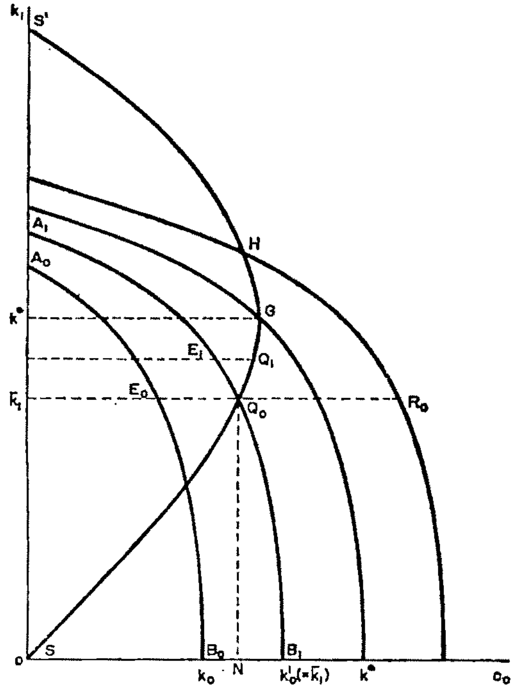


FIGURE 3

from which it follows that c' is the stationary value of consumption which corresponds to $\bar{k}_1$. Using our earlier notation, we have $c'=s(\bar{k}_1)$.[14]

Let us now introduce in Figure 3 the (stationary state) function $c=s(k_1)$. This is represented by the SS' curve whose properties are derived from Figure 1. We can then see that if we start at E_o we may find a point (Q_o) on the next period's production frontier by the intersection of the horizontal line passing through E_o and the SS' curve. If in the next period the economy chooses E_1 as its equilibrium point, then in the following period the production frontier must pass through Q_1. A similar analysis applies to the possibility where the equilibrium point is to the right of SS', as in the case when the economy is initially at R_o. The production frontier of the next period will then pass through Q_o.

Some additional features of Figure 3 should be noted. We know from our earlier discussion that the point G corresponds to the golden rule levels of c and k. At this point the slope of the production frontier AB is unitary, i.e., $f'(k^*-c^*)=1$. Similarly, for any point (c, k) on SS' below G the slope of the production frontier is $f'(k-c)>1$, and above G we have $f'(k-c)<1$. In fact, the slope (in numerical value) of the production frontiers decreases steadily as we move upward along the SS' curve. This is so since as we move up the SS' curve we increase steadily the value of k_1 and hence of k_o-c_o (via $k_1=f(k_o-c_o)$) so that $f'(k-c)$ decreases. Another point which should be noted is that at any point on SS' above G (such as H), the slope of SS' itself is always steeper, i.e., greater in numerical value, than the slope of the AB curve which passes through that point. This can be seen by differentiating the equation $k=f(k-c)$ which determines the SS' curve. This yields

$$\frac{dk}{dc}=\frac{-f'}{1-f'}\left[=\frac{1}{s'(k)}\right]. \tag{30}$$

Since above G we have $0<f'<1$,[15] it follows that $|dk/dc|>|-f'|$ as stated. Finally, let the equation of OG, along SS', be denoted by $k=y(c)$, so that $y(c)$ is the inverse of $s(k)$ for $k<k^*$. Then as c tends to zero, $y(c)$ tends to zero, and consequently, $f'(k-c)$ tends to infinity by (10'). It follows therefore from (30) that $y'(0)=1$.

The optimal path can now be easily determined if we introduce the income-consumption curve EE' into the picture, as we do in Figure 4. Denote the equation corresponding to EE' by $k=e(c)$. Then from our foregoing analysis we know that $e'>0$. Suppose that at the origin we have $e'(0)>y'(0)$ $(=1)$.[16] Then because of $e'>0$

[14] If $\bar{k}_1>\hat{k}$ (see Figure 1), then $c'=s(\bar{k}_1)<0$, implying that we have to use the extended part of SS' in Figure 3 which continues to the left of the k_1 axis. Clearly, in this case, c' does not represent actually maintainable consumption. It does, however, serve the purpose of determining graphically the next period's production frontier, as in the case with $c'>0$.

[15] In fact, by (9') we always have $f'\geq 1/n>0$.

[16] In the logarithmic model, where $e(c_o)=Ac_o^{\gamma}$ we have $e'(0)=+\infty$.

we must have an intersection of EE' and SS' for some positive k and c. Let M be an intersection point of the two curves. Since this point is both a stationary solution and a short-run equilibrium, it must represent a stationary optimum. We have seen earlier that in a stationary optimum $f'(\bar{k}-\bar{c})=1/a$ where $\bar{k}$ and $\bar{c}$ are determined by (18) and (19). It follows therefore that the intersection of EE' and SS' is unique and that the intersection point must be below G,[17] as drawn in Figure 4. Suppose alternatively that at the origin $e'(0)<1$. Then, since a (unique) positive stationary optimum is known to exist, the EE' curve must intersect[18] SS' from below at some positive k. However, this implies an additional intersection, which contradicts the uniqueness of the stationary optimum solution.

Suppose now that the economy starts with some k_o. Then the short-run equilibrium is determined in Figure 4 by the intersection of A_oB_o and EE' at the point E_o. The next period's production frontier is then A_1B_1 which passes through S_o and the new short-run equilibrium is at E_1. It can then be seen that the system converges monotonically and asymptotically to the (stationary optimum) point M from below. Similarly, had we started with $k_o>\bar{k}$, the system would converge to M from above.

Consider the golden rule point G. Since this point is necessarily above EE', it follows that at G the marginal rate of substitution of k_1 for c_o on the consumption side is greater than on the production side, i.e., c_o is more valuable (at the margin) in consumption than in production. Thus if the economy is given an initial capital

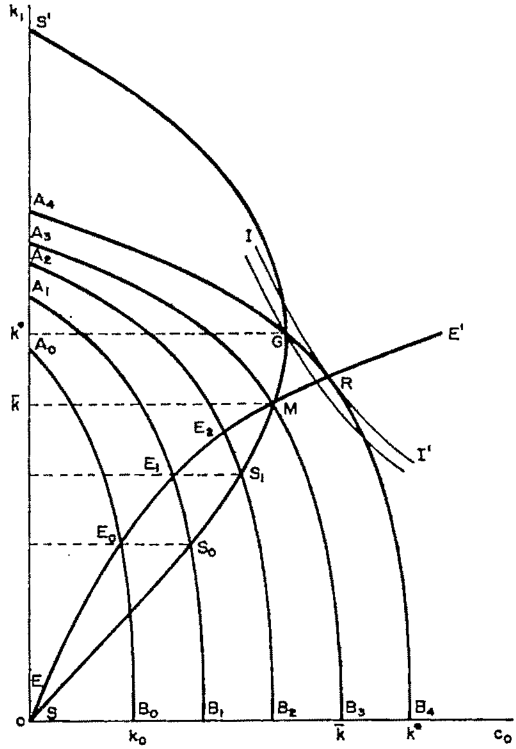


FIGURE 4

equal to k^*, it will not stay at G but will rather increase consumption immediately to the point R, and then, in subsequent periods, will reduce consumption and capital gradually to the point M.

[17] Since $1/a=f'(\bar{k}-\bar{c})>1$.

[18] The possibility of tangency can be ruled out by simple considerations of continuity. Thus suppose that EE' is below SS' except at M, where the two curves are tangent. Then for initial $k_o>\bar{k}$ the sequence c_t will converge to $\bar{c}$, while for $k_o<\bar{k}$ c_t will converge to zero. This implies that $v(k_o)$ is discontinuous at $\bar{k}$, contrary to our assumptions.

REFERENCES

R. Bellman, *Dynamic Programming*, Princeton 1957, pp. 11-16.

D. Cass, "Optimum Growth in an Aggregative Model of Capital Accumulation," *Rev. Econ. Stud.*, July 1965, *32*, 233-40.

T. C. Koopmans, "On the Concept of Optimal Economic Growth," *Pontificia Academia Scientiarus*, Rome 1965, pp. 276-79.

E. S. Phelps, "The Golden Rule of Accumulation: A Fable for Growthmen," *Amer. Econ. Rev.*, Sept. 1961, *51*, 638-42.

R. Radner, *Notes on the Theory of Economic Planning*, Athens 1963.

R. H. Strotz, "Myopia and Inconsistency in Dynamic Utility Maximization," *Rev. Econ. Stud.*, 1956, No. 3, *23*, 165-80.

Marginal Cost Pricing of Airport Runway Capacity

By ALAN CARLIN AND R. E. PARK*

In an increasing number of *U.S.* cities the rapidly growing demand for use of airport runways experienced in the last decade has exceeded the available capacity, resulting in intolerably long delays to aircraft attempting to use them. Although the problem has been most acute and has received the most publicity in New York, similar problems have already arisen in Chicago and Washington and are likely to arise elsewhere in the years to come as the demand for air travel increases.

When faced with such a problem, most economists are likely to consider first the possibility of imposing marginal cost pricing or congestion tolls as a means to optimize the use of the given transportation facilities.[1] It is primarily this approach that we will explore here. Although other short-term solutions to the congestion problem are both feasible and interesting, they will be only briefly considered.

Throughout the paper our primary example will be New York's LaGuardia Airport during the period April 1967 through March 1968.[2] This is the most recent period for which data were available at the time the research was carried out.

The paper is divided into two sections. The first develops marginal delay or congestion costs for LaGuardia. The second explores the possible use of congestion tolls as a solution to the short-term congestion problem.

I. *Marginal Delay Costs*

What are the congestion costs that an additional user would impose on others? Equivalently, what would the savings to others be if one fewer plane were to use LaGuardia? In order to answer these questions, it is necessary to realize that during a period when an airport is continuously busy, each user imposes some delay on all following users until the end of the busy period. That is, an additional user shoves those following him one space back in the queue, and the effect persists until the queue dissipates.

* The authors, economists at The RAND Corporation, are indebted to the referee for helpful suggestions. This paper is largely based on research undertaken for the Port of New York Authority, and reported more fully in our RAND memorandum. Any views expressed in this paper are those of the authors. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its research sponsors.

[1] See, for example, M. Beckmann, C. B. McGuire, and C. B. Winsten, R. H. Strotz, and W. Vickrey. For a simple model that emphasizes some features of air transport, see Park. The rationale for congestion tolls is simple: A marginal user imposes congestion costs on other users, but considers only those costs that he bears himself in deciding to use a facility. A charge equal to the marginal congestion cost would internalize the external cost, so that individual decisions would lead to a socially efficient balance between use and congestion. This is clearly a partial argument. To the extent that there are offsetting external benefits from use of the facility, or that alternatives to use also involve external costs, the argument for a congestion toll is weakened. Indeed, any uncorrected departure from optimal conditions elsewhere in the economy makes it unlikely that a partially optimal congestion toll will result in a global second-best allocation. But these considerations do not rob the notion of a congestion toll of its merit. In many real cases, such as the one examined in this paper, it may be possible to say with considerable confidence that use of the facility with a congestion toll, although not globally optimal, would be more efficient than use without the toll.

[2] A similar analysis for Kennedy International Airport in New York is reported in Carlin and Park.

The Model

We seek the delay costs, C_i, imposed by a user of type i on other users at a time t when the remaining busy period equals B minutes. A type of use is defined by the type of plane and by specification as to whether it is a landing or takeoff. Thus $i=1$ may be a large jet landing, $i=4$ a light plane takeoff, and so on. Say there are m different types of use. Then if we knew the absolute service times, S_i, the number of operations of each type, N_i, that would occur from time t until the end of the busy period, and the costs per minute of delay to each type of operation, c_i, it would be easy to calculate the marginal delay costs due to an additional operation of type i as

$$C_i = S_i \sum_{i=1}^{m} N_i c_i \tag{1}$$

The operation delays each of the $N = \sum_{i=1}^{m} N_i$ operations for the length of time it takes to service it, S_i, at a cost to each of c_i per minute. But the N_i and the S_i are awkward to estimate. It is somewhat easier to estimate relative service times $s_i = S_i/S_1$, and proportions of various types of operations $n_i = N_i/N$. So we shall transform the relationship into a more usable form by introducing a second relationship. The length of the remaining busy period must just equal the sum of the time necessary to service each of the airplanes that lands or takes off before it ends:

$$B = \sum_{i=1}^{m} N_i S_i \tag{2}$$

Dividing (1) by (2), we obtain

$$\frac{C_i}{B} = \frac{S_i \sum_{i=1}^{m} N_i c_i}{\sum_{i=1}^{m} N_i S_i} \tag{3}$$

If we divide both the numerator and denominator of the right-hand expression in (3) by $S_1 N$, we obtain

$$\frac{C_i}{B} = s_i \cdot \frac{\sum_{i=1}^{m} n_i c_i}{\sum_{i=1}^{m} n_i s_i}, \tag{4}$$

which conveniently expresses marginal cost per minute of the remaining busy period in terms of use proportions, relative service times, and individual costs per minute of delay.

At LaGuardia, which always operates with either a single or an intersecting runway configuration, arrivals and departures may be considered to be interdependent. In this case, we can approximate reality reasonably closely by distinguishing among four kinds of operations:

1—air carrier landings,
2—air carrier takeoffs,
3—general aviation landings, and
4—general aviation takeoffs.[3]

Since $s_1 = 1$, the formula for the marginal delay cost C_1 due to an air carrier landing at time t is

$$C_1 = \frac{n_1 c_1 + n_2 c_2 + n_3 c_3 + n_4 c_4}{n_1 + n_2 s_2 + n_3 s_3 + n_4 s_4} \cdot B(t) \tag{5}$$

and the costs C_i due to other types of operations are $s_i C_1$.

Empirical Estimates

The values of the n's vary throughout the day; hourly average general aviation as a percentage of total traffic on duty runways at LaGuardia varied from under 30 to over 40 percent during the busy afternoon hours.[4] So, of course, do those of the s's, in response to changing traffic mixes

[3] General aviation, consisting primarily of relatively small aircraft, is a category that includes air taxis, business and private planes.

[4] From an analysis of Federal Aviation Administration Runway Use Logs for 23 random sample days in 1967.

TABLE 1—PARAMETER ESTIMATES FOR LAGUARDIA

	Air Carrier Landings $i=1$	Air Carrier Takeoffs $i=2$	General Aviation Landings $i=3$	General Aviation Takeoffs $i=4$
1. Proportion of total traffic on duty runways, n_i	.32	.32	.18	.18
2. Service time relative to air carrier landings, s_i	1.00	.86	.54	.46
3a. Cost to aircraft owners (dollars per minute)	6.50	2.60	1.00	.50
3b. Passengers per operation	46.8	46.8	1.8	1.8
3c. Cost of passenger time (dollars per minute)	4.68	4.68	.36	.36
3d. Marginal cost of delays (dollars per minute), c_i	11.18	7.28	1.36	.86
4. Marginal cost of delays per minute of remaining busy period (dollars), $C_i/B(t)$	8.15	7.01	4.40	3.75

Notes on line:

1. From Carlin and Park (Table 5.1).
2. From Carlin and Park (Appendix Table C.3).

3a. From Carlin and Park (Appendix Table D.1), which also gives some cost figures for other airlines.
3b. Based on information supplied by Aviation Department, Port of New York Authority. General aviation and air carrier data are for 1965 and 1967, respectively.
3c. Assumes costs of $6 and $12 for air carrier and general aviation passenger time, respectively.
3d. Sum of lines 3a and 3c.

4. Computed using equation (5) and lines 1, 2, and 3d.

and runway configurations. And so, finally do those of the c's, primarily as a result of changing load factors throughout the day. However, to deal explicitly with all of these complexities would not, we feel, add enough precision to our estimates to be worth the large increase in computation that would be required. We thus limit our task in the next paragraphs to obtaining estimates of yearly *average* values for the quantities, other than $B(t)$, that enter expression (5) for marginal delay costs. Average values of $B(t)$ are estimated for each hour of the day.

Traffic proportions, n_i: Estimates of the n's are easily obtained from aggregate traffic statistics available for 1967, corrected to eliminate that fraction of general aviation traffic that used non-duty runways. These estimates are shown in line 1 of Table 1.[5]

Relative service times, s_i: By using the airport capacity manual prepared by Airborne Instruments Laboratory (*AIL*), it is possible to derive approximate ratios of the service times required by different classes of aircraft at LaGuardia.[6] Line 2 of Table 1 summarizes our estimates of relative service times.

Cost of delay to airplane owners and passengers, c_i: Line 3 shows our estimates of (or assumptions about) average cost of one minute delay to different kinds of operations, including both costs to the airplane owners and costs of passenger time. The estimates of air carrier costs are based on American Airlines figures for the types of planes that they operate at LaGuardia. In addition to direct variable costs of fuel, oil, and crew time, they include some allowance for indirect variable costs such as maintenance and incremental capital costs.[7] The fairly high general aviation costs reflect the fact that the average general aviation plane operating at La-

[5] For details, see Carlin and Park (p. 91).

[6] For details, see Carlin and Park (pp. 197–201). J. V. Yance performed a similar analysis for Washington National.

[7] For further details, see Carlin and Park (pp. 202–206). This also shows costs (one higher and one lower) reported by two other airlines.

Guardia is quite sophisticated. We assume $6 per hour as an average value for airline passenger time, and $12 per hour for presumably more affluent general aviation passengers.[8]

Marginal delay costs per minute of remaining busy period, $C_i/B(t)$: When the traffic proportions, the relative service times, and the delay costs per minute shown in lines 1 through 3 of Table 1 are substituted in expression (5) for marginal delay costs, the results are as shown in line 4 of that Table. These are estimates of the marginal delay costs per minute of remaining busy period that incremental operations of different types impose on other users.

Average remaining busy period, $B(t)$: To complete our estimates of marginal delay costs, we need information on the average length of remaining busy period by time of day, $B(t)$. As far as we know, this is the first time that an attempt has been made to estimate remaining busy periods. The method that we used is in principle a simple one. The basic data, kindly provided by American Airlines and United Air Lines, relate to delays experienced by individual flights. These data can be used to block out periods during which the airport was busy. For example, if an airplane that took off at 1615 was delayed for 15 minutes, this would ordinarily be an indication that the airport was busy between 1600 and 1615 on that day.[9] After busy periods were blocked out, the length of busy period remaining was tabulated at 10 minute intervals from 0700 to 2400 for each of 14 sample days and during 6 critical afternoon hours for each of 14 additional sample days. These values were averaged over all sample days to estimate expected busy period remaining by time of day. The estimates, together with standard errors, are shown in an appendix to this article which the authors will provide upon request. Hourly averages of the every-ten-minute estimates are shown in column 1 of Table 2.

Full marginal delay costs: It now remains only to multiply the costs per minute of remaining busy period shown in line 4 of Table 1 by the busy period estimates presented in column 1 of Table 2. The result-

[8] For comparison, one study finds a value of $2.82 per hour for commuting motorists by a study of their behavior. See T. C. Thomas. Estimated marginal costs are sensitive to this assumption. Estimates for higher and lower values of passenger time are given in fn 10.

[9] In practice, there were a number of complications. In the first place, we had to use a number of different methods to calculate the delays experienced by individual flights. For some American arrivals, we had pilot reports of delays enroute and in the New York terminal area, both of which can in large part be attributed to airplane congestion at LaGuardia. For such flights we used the pilot reports as the delay measure. For some American arrivals, pilot reports were missing. In such cases, we based the delay estimate on the excess of actual over planned flight time, a measure that reflects other influences, such as enroute wind and weather forecast errors, in addition to terminal area delays. A random sample of about 50 flights for which both pilot-reported delays (*PRD*) and excess of actual over planned flight time (*EFT*) were recorded showed the following regression relationship:

$$PRD = .05 + \underset{(.04)}{.92}\, EFT, \qquad R^2 = .90,$$

which we used to estimate delays when pilot reports were missing. For a few American flights, and all United flights, neither *PRD* nor *EFT* were available; there are no satisfactory delay measures in such cases. For all departures, both American and United, we calculated delays as the actual time elapsed between gate departure and takeoff, less a standard taxi time from the terminal building to the takeoff runway in use.

A second complication is that the data were occasionally contradictory. For example, one flight took off at 1615 after a calculated delay of 15 minutes, while another took off at 1610 on the same day with no delay. Much more frequently, there were gaps in the observations on individual flights. Although the observations were dense enough during some times of day to show several planes waiting for takeoff or for landing all at the same time, at other times there were periods during which no American or United flights were either waiting or operating undelayed. To resolve contradictions and fill in gaps, we made use of Federal Aviation Administration Runway Use Logs, which record the time of landing or takeoff of all planes using LaGuardia to the nearest minute, and thus provide a rough indication of whether the airport was busy or not at any particular time.

TABLE 2—AVERAGE REMAINING BUSY PERIOD AND FULL MARGINAL DELAY COSTS BY HOUR OF DAY

Hour of Day	Remaining Busy Period (minutes)	Full Marginal Delay Costs ($ per incremental operation)[a]			
		Air Carrier		General Aviation	
		Arrivals	Departures	Arrivals	Departures
	(1)	(2)	(3)	(4)	(5)
0000–0700	0.0[b]	0[b]	0[b]	0[b]	0[b]
0700–0800	7.4	60	52	32	28
0800–0900	33.1	270	232	146	124
0900–1000	33.2	271	233	146	125
1000–1100	19.9	162	140	88	75
1100–1200	11.4	93	80	50	43
1200–1300	30.1	245	211	132	113
1300–1400	72.9	594	511	321	173
1400–1500	85.2	694	597	375	319
1500–1600	133.7	1090	937	588	501
1600–1700	118.2	963	829	520	443
1700–1800	96.4	786	676	424	361
1800–1900	74.5	607	522	328	279
1900–2000	44.6	364	313	196	167
2000–2100	19.5	159	137	86	73
2100–2200	7.2	59	50	32	27
2200–2300	1.7	14	12	8	7
2300–2400	.4	3	3	2	2

Notes:
[a] Computed using line 4, Table 1 and column 1, this Table.
[b] Assumed to be zero.

ing figures, shown in columns 2 through 5, are average values of the delay costs imposed on other users by incremental operations at any time of day. Some of these costs are very high. For example, it appears that an additional carrier arrival between 1500 and 1600 will, on the average, impose delay costs of over $1,000 on other users. Conversely, one less arrival during this period could be expected to reduce delay to others by the same amount. Marginal delay costs for general aviation operations during the same hour are in excess of $500.[10]

[10] If carrier and general aviation passenger time were valued at $3 and $6 per hour, respectively, the costs shown in Table 2 would be reduced by 25 percent. For $12 and $24 per hour, the costs would be increased by 49 percent.

There are apparently substantial savings to be realized by some reduction in low-value traffic at LaGuardia. Some approaches to achieving a reduction are discussed in the next section.

II. *Policy Alternatives*

No Change in Policy

During the study year, flight fees at LaGuardia were based on airplane weight, with a $5 minimum for each takeoff and no charge for landing. Most general aviation pays the minimum fee. Fees for carriers range from about $50 to $150, depending on the weight of the plane. Similar value-of-service fees are used at almost all major airports.

This fee structure leads to two related

inefficiencies. First, there is an inefficiently large amount of general aviation traffic. As shown in Table 1, 36 percent of all duty runway traffic at LaGuardia during 1967 was general aviation. Since most paid only \$5 per landing and takeoff, and marginal congestion costs run up to on the order of 200 times that amount, one must conclude that many general aviation operations are of very low value relative to the congestion costs they impose on others.

It should be pointed out that on August 1, 1968, the Port of New York Authority raised the minimum fee to \$25 for flights that land or take off between 0800 and 1000 Monday through Friday, and between 1500 and 2000 every day. This pioneering but limited step in the direction of marginal congestion cost pricing appears to have reduced significantly the amount of general aviation traffic at LaGuardia. The limited information available suggests that general aviation may have been reduced as much as 40 percent at LaGuardia during the hours when the \$25 minimum applies.

Second, airline passenger loads are inefficiently low. At LaGuardia, airline load factors averaged 59.4 percent during 1967.[11] In the airline industry, competition is primarily on the basis of service rather than price.[12] One very important part of this service competition is competitive scheduling. With the present level of fares, costs are covered at fairly low load factors. The airlines tend to add flights to the same destination at roughly the same time until load factors are forced down toward these break-even levels. Although the higher frequency of service that results is not without value, it seems certain that less frequent service at higher load factors would be more efficient.

In the remainder of this section we shall discuss two pricing approaches to increasing the efficiency with which the runways are used.

Full Marginal Cost Pricing

One of the difficulties in making estimates of marginal costs with the intention of using them as prices is to allow for the effects that the prices themselves will have on runway use. Use of the costs shown in Table 2, for example, would not represent equilibrium conditions because their use would reduce the number of airplanes using the airport, thereby reducing marginal costs and hence the prices that should be charged. If instituted immediately such a pricing system would be less than optimally efficient by overly reducing traffic. It might even be dynamically unstable, in the sense that costs recomputed in succeeding periods and used as prices might not converge to an equilibrium level. This would be the case if, after Table 2 prices were imposed, traffic decreased so much that marginal costs fell below the level of current flight fees. Using these lower costs as prices during the following period would result in traffic above current levels, and undamped oscillations in prices and traffic would ensue.

Using Table 2 costs as prices without adjustment would clearly be unwise. On the other hand, to determine analytically a set of equilibrium prices would be an impossible task. To do so, we would need to know with some confidence and precision what the pattern of traffic would be under different sets of prices. We do not. One way to implement equilibrium marginal cost prices in light of these difficulties would be to charge an increasing percentage of the full marginal costs as recomputed after each successive increase.

Efficiency: Equilibrium marginal cost prices would result in very efficient runway use. They would obviously exclude low

[11] Aviation Division, Port of New York Authority.

[12] For an extended discussion, see R. E. Caves (pp. 331–55).

value general aviation traffic. They would also increase carrier load factors to a more efficient level.[13]

Practicality: Equilibrium marginal cost pricing does not appear to be a practical policy at LaGuardia. At least in the short run, airlines would be hurt by higher flight fees. The formula for calculating present airline flight fees is embodied in lease agreements between the Port Authority and the individual airlines. The airlines would therefore be able to block any move by the Port to impose higher fees.

In the longer run, there may be some offsetting considerations. For one thing, the higher flight fees would be (in principle, completely) offset by reduced operating costs as schedules were reduced and load factors increased. However, given existing airline fleet sizes and commitments for additional airplanes, it would take a long time for this adjustment to work itself out. For another thing, the increased flight fee revenues might be used by the Port Authority for capacity expansion, as Herbert Mohring suggests. However, the constraints on capacity expansion are primarily political, not financial; additional Port Authority revenue would do little or nothing to promote it.

These possible benefits are remote or uncertain enough so that short-run considerations would surely dominate. Airline opposition to full marginal cost pricing would keep it from being a practical alternative.

Proportional Marginal Cost Pricing

A more practical pricing approach is to limit total airline runway use payments to what they would be under the formulas written into the present leases, but attempt to change the basis on which fees are levied so that fees during any hour would be proportional to those that would prevail under full marginal cost pricing. Even if equilibrium marginal cost pricing were to be attempted, this would be the recommended first step toward it.

Using the full marginal costs shown in Table 2 as a basis, it is a simple step to compute what proportional fees would be on the assumption that the airlines as a group are to pay no more than they otherwise would. Using the average of September 1967 and February 1968 airline schedules, hypothetical collections with full marginal cost flight fees were computed. A percentage was then computed by dividing this number into actual airline payments to the Port Authority for runway use during the period March 1967 through February 1968. When these percentages were used to derive proportional cost fees, the prices shown in Table 3 resulted.

Practicality: Although the factors of proportionality have been chosen to keep payments by airlines as a group constant, particular airlines may pay considerably more or less under such a scheme than under the present weight-based system. Presumably the less individual airline payments exceeded present payments, the easier it would be to obtain airline agreement to such a system.

In order to explore this question, we compared what airline payments would have been using the prices shown in Table 3 and average schedule data for September 1967 and February 1968 with actual payments for the period March 1967 through February 1968. The results, shown as percentages, are presented in Table 4.

With one major exception, the calculations are fairly encouraging for airline acceptance of such a scheme at LaGuardia. The major exception is the local service airlines, particularly Allegheny and Mohawk. Because they use smaller planes which benefit from the present weight related fees, and have few flights in the

[13] For a theoretical treatment of this latter point, see Park.

TABLE 3—PROPORTIONAL MARGINAL COST PRICES FOR LAGUARDIA YIELDING CURRENT PORT AUTHORITY AIRLINE REVENUE
(dollars per operation)

Hour of Day	Air Carrier		General Aviation		Post-August 1968[a] Actual Minimums	
	Arrivals	Departures	Arrivals	Departures	Departures	Either[b]
0000–0700	0[c]	0[c]	0[c]	0[c]	5	
0700–0800	7	6	4	3	5	
0800–0900	30	26	16	14		25[d]
0900–1000	31	26	17	14		25[d]
1000–1100	18	16	10	9	5	
1100–1200	11	9	6	5	5	
1200–1300	28	24	15	13	5	
1300–1400	67	58	36	31	5	
1400–1500	78	67	42	36	5	
1500–1600	123	106	66	57		25
1600–1700	109	94	59	51		25
1700–1800	89	76	48	41		25
1800–1900	69	59	37	32		25
1900–2000	41	35	22	19		25
2000–2100	18	15	10	8	5	
2100–2200	7	6	4	3	5	
2200–2300	2	1	1	1	5	
2300–2400	0	0	0	0	5	

Notes:
[a] Prior to August 1968, the minimum was a uniform $5 throughout the day for departures only.
[b] For both arrival and departure if either occurs during the hours shown.
[c] Assumed to be zero.
[d] Monday through Friday only. The $5 minimum departure fee applies on other days.
Source:
Table 2.

early morning hours, they would be particularly affected by proportional cost fees. On the other hand, their higher runway use fees would be at least partially offset by reduced delay costs. Furthermore, airlines as a group would certainly be better off on balance. Since the group is a small one, one might expect informal pressures or side payments to bring about agreement by all airlines to proportional cost fees.

Efficiency: The proportional cost prices in Table 3 would do much to deter low value general aviation traffic during busy hours. The significant effect of a $25 charge for landing and takeoff combined is mentioned above. Proportional cost fees, which range up to almost five times this amount, would probably eliminate almost as much general aviation traffic as would full marginal cost prices. On the other hand, proportional prices by themselves would have little effect on inefficiently low load factors. Flight fees for carrier aircraft using La Guardia currently range from about $50 to $150 per landing and takeoff combined. If the proportional marginal cost prices of Table 3 were in effect, a carrier plane landing and taking off at peak times would pay as much as about $200. This increase is small enough so that it would probably

TABLE 4—CALCULATED AIRLINE PAYMENTS FOR LAGUARDIA RUNWAY USE USING PROPORTIONAL MARGINAL COST PRICING AS A PERCENTAGE OF ACTUAL COLLECTIONS

Trunk carriers	90
American	100
Eastern	66
National	56
Northeast	122
TWA	90
United	88
Local service carriers	230
Allegheny	263
Mohawk	242
Piedmont	142

have little influence on the schedules of the air carriers (with the possible exception of the local service carriers).

Other Alternatives

We have examined in some detail two pricing approaches to increasing the efficiency with which La Guardia's runways are used. Full marginal cost pricing is the most efficient pricing scheme, but it could not be adopted over airline opposition, and airline opposition is likely. Proportional marginal cost pricing is probably the most efficient pricing scheme with a reasonable chance of being implementable, but it does little to correct inefficiently low airline load factors. This suggests that other measures capable of affecting schedules as well as general aviation use would be even more efficient than proportional marginal cost pricing. These other alternatives include purely administrative measures and combination measures involving both pricing and administrative aspects. Although a complete discussion is beyond the scope of this paper, we shall mention a few of the possibilities here.[14]

The most efficient purely administrative measures would restrict both general aviation and air carriers to something less than their current levels of operation. One central problem in the design of an efficient administrative measure is to decide what level each should be restricted to. Another is to formulate the restrictions so as to attempt to exclude the lowest value traffic in each category. On June 1, 1969, the Federal Aviation Administration imposed limits on operations during bad weather at five busy airports, including LaGuardia. Although the limits are higher and allocate a larger share to general aviation than would be efficient, this measure is a small step toward more efficient runway use. However, bargaining among the airlines is a cumbersome way to split up the quota, and there is no guarantee that the schedules will go to the highest value users.

There are at least two combination measures that are potentially more efficient than any of the feasible "pure" alternatives. The first of these would consist of the issuance of property rights in schedule slots for particular hours in proportion to current use that could be freely traded among airport users.[15] The number of slots for each hour would be chosen administratively to approximate the efficient number. The free market in slots would then allocate them to the highest-value users. It would, of course, be permissible to subdivide the slots so that, for example, an airline might purchase the right to use a slot on Monday through Friday only. If the slots were issued to airlines in proportion to recent schedules, all would share in the gains from increased efficiency; there should be no airline opposition to overcome in introducing this measure. In theory, one could issue many small fractional slots to nonscheduled users as well, and rely on the market to allocate slots among *all* users. In practice, it would probably be

[14] For a more detailed discussion, see Carlin and Park (pp. 139–55).

[15] This approach was suggested by Jack Hirshleifer.

better to rely on some pricing mechanism to control nonscheduled users. As another practical matter, it would be best to issue the property rights in slots on a relatively short-term basis, say one or two years, to make it easier to adjust the number of slots or otherwise to modify the system.

The second combination measure would consist of both proportional marginal cost pricing and administrative limits on airline schedules. The proportional prices would exclude low-value general aviation users, and the schedule limits would increase airline load factors.

III. *Conclusions*

Because of the practical problems involved, equilibrium marginal cost pricing does not appear to be a feasible alternative for allocating runway capacity at La Guardia. The use of proportional marginal cost pricing, however, offers some of the same efficiency advantages without most of the problems. It is certainly preferable on efficiency grounds to the present weight-based, value of service pricing used at most airports. Use of administrative limits on schedules would be required in conjunction with proportional marginal cost pricing, however, to increase carrier load factors to more efficient levels.

REFERENCES

M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation.* New Haven 1956, pp. 80-101.

A. Carlin and R. E. Park, *The Efficient Use of Airport Runway Capacity in a Time of Scarcity,* RM-5817, The RAND Corporation. Santa Monica Aug. 1969.

R. E. Caves, *Air Transport and Its Regulators: An Industry Study.* Cambridge, Mass. 1962.

H. Mohring, "Urban Highway Investments," in R. Dorfman, ed., *Measuring Benefits of Government Investments.* Washington 1965, pp. 231-75.

R. E. Park, "Congestion Tolls for Regulated Common Carriers," P-4153, The RAND Corporation. Santa Monica Sept. 1969. Also forthcoming in *Econometrica.*

R. H. Strotz, "Urban Transportation Parables," in Julius Margolis, ed., *The Public Economy of Urban Communities.* Washington 1965.

T. C. Thomas, *The Value of Time for Passenger Cars: An Experimental Study of Commuter's Values,* Palo Alto 1967.

W. Vickrey, "Optimization of Traffic and Facilities," *J. Transp. Econ. Policy,* May 1967, *1,* 123-36.

J. V. Yance, "Movement Time as a Cost in Airport Operations," *J. Transp. Econ. Policy,* Jan. 1969, *3,* 28-36.

Airborne Instruments Laboratory *Airport Capacity,* prepared for Federal Aviation Agency and available from Center for Scientific and Technical Information as PB 181 553, New York June 1963.

Port of New York Authority, Aviation Department, *Airport Statistics.*

A Growth Model of International Direct Investment

By Hans Brems*

Few growth models incorporate international trade, fewer still, international investment,[1] and none, international direct investment. The purpose of the present paper is to build and solve a two-country four-sector equilibrium-growth model of international direct investment. Why should such investment exist? The factor-price-equalization theorem rests upon two assumptions: first, that no production function known to the entrepreneurs of one country is not known to those of the other; and second, that no good produced by the entrepreneurs of one country is not produced by those of the other.

Relaxing the twin assumptions of commonly known production functions and nonspecialization, we cannot expect factor-price equalization. Consequently incentives to factor movements exist. We assume that labor cannot move internationally at all. We define international direct investment as the movement of a bundle of money capital *plus* technological and managerial knowledge from a parent firm to its foreign subsidiary. In this form, money capital is known to flow more easily and in larger volume[2] than in the form of portfolio investment, so we assume it to be the only possible form.

An international model must have at least two countries in it. Since international direct investment generates a foreign-owned sector in the host country, each country may have two sectors in it, a domestically-owned and a foreign-owned one. We build our four-sector model as follows. In the ith country, let the jth country's entrepreneurs produce a unique good whose output is X_{ij}. In neoclassical tradition (see Robert Solow), this good is produced from, first, the ith country's labor employed by the jth country's entrepreneurs L_{ij} and, second, an immortal capital stock, S_{ij}, of that good. An act of investment, I_{ij}, by the jth country's entrepreneurs in the ith country is an act of setting aside part of their output there for installation there. If the ith and the jth country are not the same country, that act is an act of international direct investment. The jth country's entrepreneurs allocate their investment between parent firm and foreign subsidiary such as to maximize the present worth of all their future profits. Let the four commodities be

* The author is professor of economics at the University of Illinois at Urbana. He gratefully acknowledges research support from the University's Center for International Comparative Studies while on sabbatical leave in 1968 to study international direct investment in, and by, Germany and Sweden. He expresses his sincere thanks to the editor and two anonymous referees of this *Review*, to Horst Herberg (Hamburg), Lars Lund (Copenhagen), and to discussants at Aarhus, Copenhagen, Groningen, Helsinki, Kiel, Lund, Tampere and particularly the international business seminar at Uppsala.

[1] George Borts and Jerome Stein considered a one-country (one-region) open economy in which capital movements are induced by an external disturbance of its equilibrium growth path, hence are no part of a growth equilibrium. Hamada considered capital movements part of a growth equilibrium in a two-country model having only one good and one production function in it. These assumptions prevent one from distinguishing between direct and portfolio investment.

[2] At the end of 1967, the accumulated *U.S.* private direct investment abroad was 2.7 times larger than accumulated private long-term portfolio investment, see Nelson and Cutler.

good, but not perfect, substitutes in consumption, and let each country's consumers have a taste for all four of them. Let C_{hij} be the eight consumptions in the hth country of goods produced in the ith country by the jth country's entrepreneurs.

The money wage rate, w_i, specified for the ith country, is the same in the domestically-owned and the foreign-owned sector of that country. The two national money wage rates are parameters. But the four goods prices, P_{ij}, are variables, consequently the real wage rate in terms of any of the four goods is also a variable. Furthermore, the exchange rate is a variable. Such reliance on the price mechanism enables us to specify equilibrium conditions for each national labor market, each of the four goods markets, and the balance of payments. The latter includes international trade, international investment, and repatriation of profits. Figure 1 shows all physical flows and Figure 2 all money flows in our model.

Section I defines variables and param-

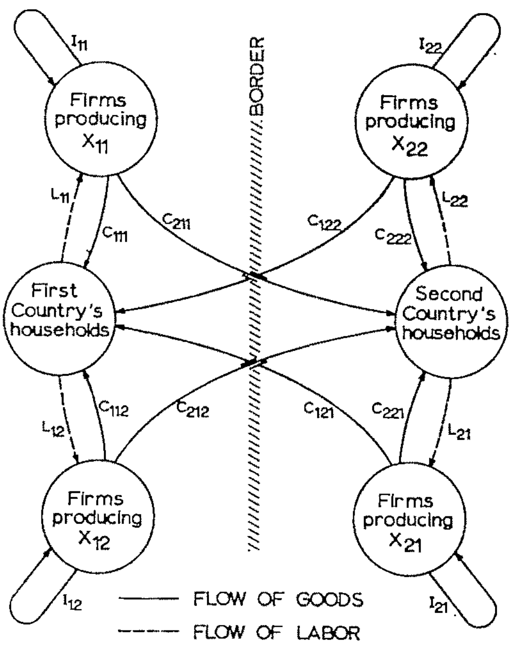


FIGURE 1. PHYSICAL FLOWS

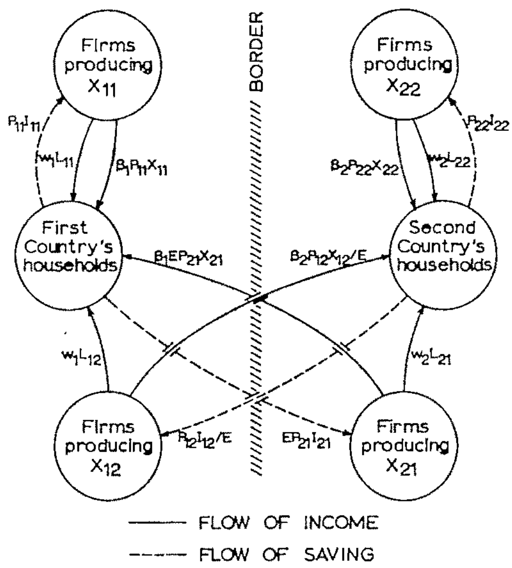


FIGURE 2. MONEY FLOWS. DIRECT INVESTMENT ABROAD AND PROFITS EARNED ON IT ARE SHOWN IN THE MONETARY UNIT OF THE INVESTING COUNTRY

eters. Section II specifies the model mathematically. Section III finds equilibrium solutions for growth rates, and Section IV finds equilibrium solutions for levels of important variables. Sections V through IX offer four numerical examples of our equilibrium solution illustrating *U.S.* direct investment abroad. Conclusions are drawn in Section X.

I. *Notation*

Variables

C_{hij} ≡ consumption in hth country of goods produced in ith country by jth country's entrepreneurs

E ≡ exchange rate in number of monetary units of Country 1 exchanged for one monetary unit of Country 2

G ≡ proportionate rate of growth of exchange rate E

$g_{C_{hij}}$ ≡ proportionate rate of growth of consumption C_{hij}

$g_{I_{ij}}$ ≡ proportionate rate of growth of investment I_{ij}

$g_{L_{ij}}$ ≡ proportionate rate of growth of employment L_{ij}

$g_{P_{ij}} \equiv$ proportionate rate of growth of price P_{ij}
$g_{S_{ij}} \equiv$ proportionate rate of growth of capital stock S_{ij}
$g_{X_{ij}} \equiv$ proportionate rate of growth of output X_{ij}
$g_{Y_i} \equiv$ proportionate rate of growth of national money income Y_i
$I_{ij} \equiv$ investment by jth country's entrepreneurs in ith country
$\kappa_{ij} \equiv$ physical marginal productivity of capital stock in ith country owned by jth country's entrepreneurs
$L_{ij} \equiv i$th country's labor employed by jth country's entrepreneurs
$P_{ij} \equiv$ price of good produced in ith country by jth country's entrepreneurs, expressed in ith country's monetary unit
$R_j \equiv$ sum of sector revenues of jth country's entrepreneurs
$S_{ij} \equiv$ capital stock of goods produced in ith country by jth country's entrepreneurs
$U_{ik} \equiv$ utility to kth person in ith country
$W_i \equiv i$th country's wage bill
$X_{ij} \equiv$ output of good produced in ith country by jth country's entrepreneurs
$Y_i \equiv i$th country's national money income
$Z_j \equiv$ current profits of jth country's entrepreneurs
$\zeta_j \equiv$ present worth of all future profits of jth country's entrepreneurs

Parameters

$A_{hij} \equiv$ parameters of individual utility function in hth country
$\alpha_j, \beta_j \equiv$ parameters of production function of jth country's entrepreneurs
$c_i \equiv i$th country's propensity to consume its national money income
$e \equiv$ Euler's number, the base of natural logarithms
$F_i \equiv i$th country's available labor force
$g_{F_i} \equiv$ proportionate rate of growth of available labor force F_i
$g_{M_j} \equiv$ proportionate rate of growth of multiplicative factor M_j
$g_{w_i} \equiv$ proportionate rate of growth of money wage rate w_i
$M_j \equiv$ multiplicative factor in production function of jth country's entrepreneurs
$N_i \equiv$ multiplicative factor in utility function in ith country
$r_j \equiv$ discount rate applied by jth country's entrepreneurs
$w_i \equiv i$th country's money wage rate

For time coordinates we shall use t and τ. The flow variables C_{hij}, I_{ij}, and X_{ij} are measured in physical units consumed, invested, or produced, respectively, per annum of the good produced in ith country by jth country's entrepreneurs. All flow variables refer to the instantaneous rate of that variable measured on a per annum basis. Subscripts h, i, and j refer to country number.

II. *The Equations of the Model*

To the eight variable growth rates listed in Section I, apply the definition of the proportionate rate of growth of a variable v

(1) through (31) $$g_v \equiv \frac{dv}{dt}\frac{1}{v}$$

Define investment as the derivative of capital stock with respect to time:

(32) through (35) $$I_{ij} \equiv \frac{dS_{ij}}{dt}$$

When producing in the ith country, let the jth country's entrepreneurs apply the Cobb-Douglas production function

(36) through (39) $$X_{ij} = M_j L_{ij}^{\alpha_j} S_{ij}^{\beta_j}$$

where $0<\alpha_j<1$, $0<\beta_j<1$, $\alpha_j+\beta_j=1$, and $M_j>0$. While the jth country's entrepreneurs do not produce the same good at home and abroad, their α_j, β_j, and M_j are the same: If disposed at home toward highly automated methods, they will be so

disposed in their foreign subsidiaries as well. Technological progress made at home will be applied in the foreign subsidiaries as well.

In any sector let profit maximization under pure competition equalize real wage rate and physical marginal productivity of labor:

(40) through (43) $$\frac{w_i}{P_{ij}} = \frac{\partial X_{ij}}{\partial L_{ij}} = \alpha_j \frac{X_{ij}}{L_{ij}}$$

Define the physical marginal productivity of capital as

(44) through (47) $$\kappa_{ij} \equiv \frac{\partial X_{ij}}{\partial S_{ij}} = \beta_j \frac{X_{ij}}{S_{ij}}$$

Define the revenue of the jth country's entrepreneurs as

(48) $$R_1 \equiv P_{11}X_{11} + EP_{21}X_{21}$$

(49) $$R_2 \equiv P_{12}X_{12}/E + P_{22}X_{22}$$

It follows from (44) through (47) applied to (48) and (49) that at time t the jth country's entrepreneurs will be earning the profits

(50), (51) $$Z_j(\mathrm{t}) = \beta_j R_j(\mathrm{t})$$

As seen from time τ these profits are

$$Z_j(\mathrm{t}, \tau) \equiv Z_j(\mathrm{t})e^{-r_j(\mathrm{t}-\tau)}$$

where r_j is the discount rate applied by the jth country's entrepreneurs. The present worth of all future profits as seen from time τ is

(52), (53) $$\zeta_j(\tau) \equiv \int_{\tau}^{\infty} Z_j(\mathrm{t})e^{-r_j(\mathrm{t}-\tau)}d\mathrm{t}$$

Let the jth country's entrepreneurs allocate their investment between parent firm and foreign subsidiary such that

(54), (55) $$\zeta_j(\tau) = \text{maximum}$$

Under full employment the ith country's available labor force must equal the sum of labor employed by the jth country's entrepreneurs operating in the ith country:

(56), (57) $$F_i = \sum_{j=1}^{2} L_{ij}$$

Define the ith country's wage bill as its money wage rate *times* its combined employment:

(58), (59) $$W_i \equiv w_i \sum_{j=1}^{2} L_{ij}$$

Define the jth country's national money income as the sum of its wage bill and the profits bill earned by its entrepreneurs:

(60), (61) $$Y_j \equiv W_j + Z_j$$

Within each country let all persons have the same utility function but let the latter differ between countries. Let the utility function of the kth person in the hth country be of Cobb-Douglas form with the exponents A_{kij}. From his utility function and budget constraint, Appendix I finds first the person's own and next his country's four Graham-type[3] demand functions

(62) $$C_{111} = \pi_{111} Y_1/P_{11}$$

(63) $$C_{112} = \pi_{112} Y_1/P_{12}$$

(64) $$C_{121} = \pi_{121} Y_1/(EP_{21})$$

(65) $$C_{122} = \pi_{122} Y_1/(EP_{22})$$

(66) $$C_{211} = \pi_{211} EY_2/P_{11}$$

(67) $$C_{212} = \pi_{212} EY_2/P_{12}$$

(68) $$C_{221} = \pi_{221} Y_2/P_{21}$$

(69) $$C_{222} = \pi_{222} Y_2/P_{22}$$

where

$$\pi_{hij} = \frac{c_h A_{kij}}{A_{h11} + A_{h12} + A_{h21} + A_{h22}}$$

[3] Graham demand functions have income and price elasticities of 1 and −1 respectively. The following estimates of income elasticity of *U.S.* imports are found in the literature: Ball and Mavwah, 0.91 and 1.6; Harberger, 1.00; Houthakker and Magee, 1.51; Kreinin, 1.27. The following estimates of price elasticity have been made: Ball and Mavwah −0.51 and −1.00; Harberger, −0.95; Houthakker and Magee, −0.54; Kreinin, −1.1. Income and price elasticities of the imports of 14 other major countries were estimated at 1.37 and −0.90, see Houthakker and Magee.

Sector output equilibrium requires the output in the ith country produced by the jth country's entrepreneurs to equal the sum of consumption, export, and investment demand for it, or inventory would either accumulate or be depleted. Thus

$$\text{(70) through (73)} \quad X_{ij} = \sum_{h=1}^{2} C_{hij} + I_{ij}$$

Balance-of-payments equilibrium requires Country 1's consumption of, and investment in, Country 2's goods *minus* Country 1's profits earned in Country 2 to equal Country 2's consumption of, and investment in, Country 1's goods *minus* Country 2's profits earned in Country 1, all being measured in the same monetary unit. If this equality were not satisfied, exchange reserves of the ith country would either accumulate or be depleted. Thus

$$\text{(74)} \quad \begin{aligned} &E(P_{21}C_{121}+P_{22}C_{122}+P_{21}I_{21}-\beta_1 P_{21}X_{21}) \\ &= P_{11}C_{211}+P_{12}C_{212}+P_{12}I_{12}-\beta_2 P_{12}X_{12} \end{aligned}$$

III. *Solutions for Proportionate Rates of Growth*

Equations (1) through (74) contain 74 variables: First, 8 of each of the 2 variables C_{hij} and $g_{C_{hij}}$. Second, 4 of each of the 11 variables $g_{I_{ij}}$, $g_{L_{ij}}$, $g_{P_{ij}}$, $g_{S_{ij}}$, $g_{X_{ij}}$, I_{ij}, κ_{ij}, L_{ij}, P_{ij}, S_{ij}, and X_{ij}. Third, 2 of each of the 6 variables g_{Y_i}, R_j, W_i, Y_j, Z_j, and ζ_j, and fourth, the 2 variables E and G.

We posit the following set of steady-state solutions for the equilibrium proportionate rates of growth:

$$\text{(75)} \quad G = g_{F_1} - g_{F_2} + g_{w_1} - g_{w_2}$$

$$\text{(76) through (83)} \quad g_{C_{hij}} = g_{X_{ij}}$$

$$\text{(84) through (87)} \quad g_{I_{ij}} = g_{X_{ij}}$$

$$\text{(88) through (91)} \quad g_{L_{ij}} = g_{F_i}$$

$$\text{(92) through (95)} \quad g_{P_{ij}} = g_{w_i} - g_{M_j}/\alpha_j$$

$$\text{(96) through (99)} \quad g_{S_{ij}} = g_{X_{ij}}$$

$$\text{(100) through (103)} \quad g_{X_{ij}} = g_{M_j}/\alpha_j + g_F$$

$$\text{(104), (105)} \quad g_{Y_i} = g_{F_i} + g_{w_i}$$

To convince himself that those are indeed solutions, the reader should take derivatives with respect to time of all equations (32) through (43), (48) and (49), and (56) through (74), then use definitions (1) through (31), insert solutions (75) through (105), and see that each equation is satisfied.

According to our solutions, is our steady-state growth balanced, or unbalanced? Define balanced growth as stationary mutual proportions between physical sector outputs X_{ij}. In (100) through (103) neither g_{M_j}/α_j nor g_{F_i} can be expected to be identical in all four sectors. For example, if *ceteris paribus* the entrepreneurs of one country had a higher g_{M_j} than those of the other country, the physical outputs offered by the technological leaders would be growing more rapidly than those offered by the laggards, and growth would be unbalanced as it is in the real world (see Swamy).

It does, however, follow from (75), (92) through (95), and (100) through (103) that when measured in Country 1's monetary unit, all four sector revenues $P_{ij}X_{ij}$ will grow at the same proportionate rate $g_{F_1}+g_{w_1}$.

IV. *Solutions for Levels*

We solve for the equilibrium levels at a particular time of the exchange rate, revenues, national incomes, marginal productivities of capital, foreign direct investments, and net national capital export. To do so, we need a numéraire for each country. According to (56) through (59), the ith country's wage bill in terms of parameters is $W_i = F_i w_i$, so let us use the wage bills for our numéraires.

Appendix II reduces the non-linear system (1) through (74) to two cubic equations (137) and (138) and finds a family of cases in which the two cubic equations collapse into a linear one in one variable: Assume that in the consumer's utility

function used in Appendix I, the exponents A_{kij} of the consumptions C_{kij} of the two goods produced by the jth country's entrepreneurs are the same, then in the demand equations (62) through (69):

$$\pi_{111} = \pi_{121} \tag{106}$$

$$\pi_{112} = \pi_{122} \tag{107}$$

$$\pi_{211} = \pi_{221} \tag{108}$$

$$\pi_{212} = \pi_{222} \tag{109}$$

The two goods produced by the jth country's entrepreneurs have, then, production functions with the same parameters α_j, β_j, and M_j and are raised to the same power A_{kij} in the utility function of the kth person in the ith country. Does that make them identical goods? Does our four-good model now collapse into a two-good model? No, for the two goods are still not perfect substitutes and may, therefore, sell side by side yet have different prices (they are likely to have, because they are produced in two different countries between which there is no factor-price equalization, hence no factor-proportion equalization either).

We can live with the four assumptions (106) through (109) if our main interest is the relative success of one country's entrepreneurs vis-à-vis those of the other. And linearity would enable us to find the following unique equilibrium solutions for the levels of the exchange rate, revenues, national incomes, marginal productivities of capital, foreign direct investments, and net national capital export. Let the abbreviations m_i and q_i everywhere stand for the groupings of parameters listed in Section 4 of Appendix II. First, as shown in Appendix II, find the exchange rate

$$E = W_1/W_2 \tag{110}$$

Insert (110) into (132) and (133) in Appendix II and find the combined sector revenues

$$R_1 = (q_1 + q_2)W_1 \tag{111}$$

$$ER_2 = (q_3 + q_4)W_1 \tag{112}$$

Insert (106) through (109), (111) and (112) into (128) through (131) in Appendix II and find the four sector revenues collapsing into

$$P_{11}X_{11} = (q_1 + q_2)W_1/2 \tag{113}$$

$$P_{12}X_{12} = (q_3 + q_4)W_1/2 \tag{114}$$

$$EP_{21}X_{21} = (q_1 + q_2)W_1/2 \tag{115}$$

$$EP_{22}X_{22} = (q_3 + q_4)W_1/2 \tag{116}$$

Insert (50) and (51), (111) and (112) into (60) and (61) and find the national money incomes

$$Y_1 = [1 + \beta_1(q_1 + q_2)]W_1 \tag{117}$$

$$EY_2 = [1 + \beta_2(q_3 + q_4)]W_1 \tag{118}$$

Insert (106) through (110) into (139) through (142) in Appendix II and find the four marginal productivities of capital

$$\kappa_{11} = \frac{\beta_1 g s_{11}}{1 - 2(m_1 + m_2)/(q_1 + q_2)} \tag{119}$$

$$\kappa_{12} = \frac{\beta_2 g s_{12}}{1 - 2(m_3 + m_4)/(q_3 + q_4)} \tag{120}$$

$$\kappa_{21} = \frac{\beta_1 g s_{21}}{1 - 2(m_1 + m_2)/(q_1 + q_2)} \tag{121}$$

$$\kappa_{22} = \frac{\beta_2 g s_{22}}{1 - 2(m_3 + m_4)/(q_3 + q_4)} \tag{122}$$

Finally insert (114), (115), (120), and (121) into (143), (144), and (147) in Appendix II and find Country 1's net national capital export to be the difference between the two foreign investments $EP_{21}I_{21} - P_{12}I_{12}$, where

$$P_{12}I_{12} = [q_3 + q_4 - 2(m_3 + m_4)]W_1/2$$

$$EP_{21}I_{21} = [q_1 + q_2 - 2(m_1 + m_2)]W_1/2$$

V. *Numerical Examples, Master Case*

To illustrate *U.S.* direct investment abroad, Table 1 shows four numerical examples of our equilibrium solutions. The first example is a master case of complete

TABLE 1—REVENUES, NATIONAL INCOMES, MARGINAL PRODUCTIVITIES OF CAPITAL, FOREIGN DIRECT INVESTMENTS, AND NET NATIONAL CAPITAL EXPORT

Variable		Master Case	Changed Thriftiness Case	Changed Technology Case	Changed Preference Case
Revenues	R_1	$1.33W_1$	$1.39W_1$	$1.42W_1$	$1.94W_1$
	ER_2	$1.33W_1$	$1.27W_1$	$1.41W_1$	$0.72W_1$
National Incomes	Y_1	$1.33W_1$	$1.35W_1$	$1.47W_1$	$1.49W_1$
	EY_2	$1.33W_1$	$1.32W_1$	$1.35W_1$	$1.18W_1$
Marginal Productivities	κ_{i1}	$2.25g_{S_{i1}}$	$1.29g_{S_{i1}}$	$2.89g_{S_{i1}}$	$2.94g_{S_{i1}}$
	κ_{i2}	$2.25g_{S_{i2}}$	$2.17g_{S_{i2}}$	$2.34g_{S_{i2}}$	$1.38g_{S_{i2}}$
Foreign Direct Investments	$EP_{21}I_{21}$	$0.074W_1$	$0.135W_1$	$0.082W_1$	$0.083W_1$
	$P_{12}I_{12}$	$0.074W_1$	$0.073W_1$	$0.075W_1$	$0.065W_1$
Net National Capital Export		0	$0.062W_1$	$0.007W_1$	$0.017W_1$

symmetry, hence zero net national capital export, between the two countries, using realistic values of the technology parameters and the propensity to consume, i.e., $\alpha_j=3/4$, $\beta_j=1/4$, $c_i=8/9$, and $\pi_{hij}=2/9$. Let us now modify our master case.

VI. *Changed-Thriftiness Case*

U.S. superiority is sometimes sought in capital abundance. So without changing our master-case technology and preferences let us reduce the propensity to consume of Country 1. For example, assume $\alpha_j=3/4$, $\beta_j=1/4$, $c_1=4/5$, $c_2=8/9$, $\pi_{111}=\pi_{112}=\pi_{121}=\pi_{122}=1/5$, and $\pi_{211}=\pi_{212}=\pi_{221}=\pi_{222}=2/9$. Table 1 shows that this will reduce the marginal productivities of capital earned by the entrepreneurs of both countries—although in our model, the savings of a *U.S.* saver never gets inside a European owned firm. It does, however, get inside *U.S.* owned firms competing with European owned firms everywhere (the four goods are good, but not perfect, substitutes). Because of its higher thriftiness, Country 1 now has a positive net national capital export.

VII. *Changed-Technology Case*

U.S. superiority is sometimes sought in its knowledge of highly automated production methods. So, without changing our master-case preferences and propensity to consume, let us make the technology known to the entrepreneurs of Country 1 more automated. For example, assume $\alpha_1=2/3$, $\beta_1=1/3$, $\alpha_2=3/4$, $\beta_2=1/4$, $c_i=8/9$, and $\pi_{hij}=2/9$. Table 1 shows that this will raise all variables of both countries. The marginal productivities of capital earned by the first country's entrepreneurs κ_{i1} are raised for two reasons; first, their coefficient is raised from 2.25 to 2.89, and second, according to (96) through (103) $g_{S_{i1}}$ rises with falling α_1. Because of the superior knowledge of its entrepreneurs, Country 1 again has a positive net national capital export.

VIII. *A Widening Technology Gap*

A technology gap is said to be opening up between the United States and Western Europe, so without changing our master-case exponents α_j and β_j, propensity to consume, or preferences, let us assume such a gap to be developing in the form that $g_{M1}>g_{M2}$. Such a gap can affect neither the revenues R_j, the national incomes Y_i, nor the net capital export $EP_{21}I_{21}-P_{12}I_{12}$, for g_{M_j} is absent from the solutions for those variables. But, according to (96) through (103), g_{M_j} is indeed present in the solution for $g_{S_{ij}}$ and hence in the solutions (119) through (122) for the marginal productivities of capital κ_{ij}. Consequently the technological leaders will, *ceteris paribus*, have higher marginal productivities of capital than the laggards.

IX. *Changed-Preference Case*

U.S. superiority is often sought in management rather than in capital abundance or technology. So, at unchanged master-case technology and propensity to consume, let the entrepreneurs of Country 1 have succeeded better than those of Country 2 in conceiving and developing the optimal products. As a result, the preferences for the two goods produced by the entrepreneurs of Country 1 are up, while the preferences for the other two goods are down. For example, assume $\alpha_j=3/4$, $\beta_j=1/4$, $c_i=8/9$, $\pi_{111}=\pi_{121}=\pi_{211}=\pi_{221}=3/9$, and $\pi_{112}=\pi_{122}=\pi_{212}=\pi_{222}=1/9$. Table 1 shows that this will raise the marginal productivities earned by the first country's entrepreneurs and reduce those earned by the second country's entrepreneurs. Country 1 again has a positive net national capital export.

X. *Conclusions*

We have built and solved a four-sector growth model of international direct investment. Define steady-state growth as stationary proportionate rates of growth, and balanced growth as stationary mutual proportions between physical sector outputs X_{ij}. Unlike most growth models, our model realistically (see Swamy) generates steady state but unbalanced growth: The physical outputs of the four sectors may well be growing at four different proportionate rates. If there is a ***défi américain*** in our model, this is it.

Capital movements in the form of two-way international direct investment are a permanent, realistic,[4] and nonparadoxical feature of our full equilibrium. One country's entrepreneurs may simply be superior in computers, those of the other in steel. Once our two-way international direct investment flows are determined, so is their net. The direction of the net depends, of course, upon the entire structure of the model. To illustrate this, we first offered a numerical example of complete symmetry, and hence zero net national capital export, between the two countries. We then broke the symmetry by assuming Country 1 superior in terms of either thriftiness, technology, or consumer preference for the products offered by its entrepreneurs. In all three cases, its net national capital export changed from zero to positive, and in this sense thriftiness, technology, or management each may contribute to an explanation of *U.S.* direct investment abroad.

[4] Measured in billions of dollars per annum, the outward and inward flows of direct investment were for the United States during 1967, 4.6 and 0.9, respectively (see Nelson and Cutler, p. 22), for West Germany during 1967, 0.25 and 0.70, respectively (see Deutsche Bundesbank, p. 64), and for Sweden on an average 1961–1965, 0.068 and 0.056, respectively (see Lund, p. 135).

APPENDIX I.

The Derivation of Consumption Demand Functions

Within each country, let all persons have the same utility function, but, let the latter differ between countries. Let the utility function of the kth person be

$$U_{1k} = N_1 C_{111k}^{A_{111}} C_{112k}^{A_{112}} C_{121k}^{A_{121}} C_{122k}^{A_{122}}$$
$$U_{2k} = N_2 C_{211k}^{A_{211}} C_{212k}^{A_{212}} C_{221k}^{A_{221}} C_{222k}^{A_{222}}$$

where $0<A_{hij}<1$ and $N_i>0$. In the ith country, let there be s_i persons, and let the kth person's money income be Y_{ik} where

$$\sum_{k=1}^{s_i} Y_{ik} = Y_i$$

Let all persons in the ith country spend the fraction c_i, where $0<c_i<1$, of their money income. Then the budget constraint of the kth person in the two countries is

$$c_1 Y_{1k} = P_{11}C_{111k} + P_{12}C_{112k} + E(P_{21}C_{121k} + P_{22}C_{122k})$$
$$c_2 Y_{2k} = (P_{11}C_{211k} + P_{12}C_{212k})/E + P_{21}C_{221k} + P_{22}C_{222k}$$

Maximize the kth person's utility subject to his budget constraint and find his four

demand functions. For each country add the s_i individual demand functions for each good and find (62) through (69).

Appendix II

The Mechanics of Solving for Equilibrium Levels

1. *A Country's Saving Equals Direct Investment by its Entrepreneurs*

Multiply each of the sector-output equilibrium conditions (70) through (73) by its price P_{ij} expressed in the same country's monetary unit for all four of them. Add the conditions for $ij=11$, 12 and add (74) to their sum. Add the conditions for $ij=21$, 22 and add (74) to their sum. Use (40) through (43), (48), (49), and (56) through (69) and find:

$$(123) \qquad (1-c_1)Y_1 = P_{11}I_{11} + EP_{21}I_{21}$$

$$(124) \qquad (1-c_2)EY_2 = P_{12}I_{12} + EP_{22}I_{22}$$

Are (123) and (124) trivial? Clearly not, for in general an open economy's saving equals the sum of the direct investment by its entrepreneurs in parent firm and foreign subsidiary *plus* the net change in the country's exchange reserves. But, thanks to our equilibrium condition (74), that net change is zero.

2. *Allocating Investment between Parent Firm and Foreign Subsidiary*

Using (48) through (53), (75), (92) through (95), and (100) through (103) and assuming that

$$g_{P11} + g_{X11} < r$$

$$G + g_{P21} + g_{X21} < r$$

we find the present worth of all future combined profits to the first country's entrepreneurs as seen from time τ to be

$$(125) \qquad \zeta_1(\tau) = \frac{\beta_1 P_{11}(\tau)X_{11}(\tau)}{r_1 - (g_{P_{11}} + g_{X_{11}})} + \frac{\beta_1 E(\tau)P_{21}(\tau)X_{21}(\tau)}{r_1 - (G + g_{P_{21}} + g_{X_{21}})}$$

Use (1) through (35) to write $I_{ij} = g_{S_{ij}}S_{ij}$, insert that into (123), and write the latter for time τ as

$$(123a) \quad S_{21}(\tau) = \frac{(1-c_1)Y_1(\tau) - g_{S_{11}}P_{11}(\tau)S_{11}(\tau)}{g_{S_{21}}E(\tau)P_{21}(\tau)}$$

Insert (123a) into the production functions (36) through (39) written for time τ, and insert them into (125). By now the only variables appearing in our expression for $\zeta_1(\tau)$ are E, G, $g_{P_{ij}}$, $g_{S_{ij}}$, $g_{X_{ij}}$, L_{ij}, P_{ij}, S_{11}, and Y_i. With the exception of L_{ij} and S_{11}, all these variables are assumed to be beyond the control of, but correctly foreseen by, the first country's purely competitive entrepreneurs. Maximize $\zeta_1(\tau)$ by taking the partial derivative $\partial\zeta_1(\tau)/\partial S_{11}(\tau)$ and setting it equal to zero, thus arriving at a first-order maximum condition. Use (75), (92) through (95), and (100) through (103) to see that

$$g_{P_{11}} + g_{X_{11}} = G + g_{P_{21}} + g_{X_{21}}$$

Use this, (36) through (39), (44) through (47), and (123a) to write the first-order condition as[5] $g_{S_{11}}/g_{S_{21}} = \kappa_{11}/\kappa_{21}$ or, by again using (1) through (35) to write $I_{ij} = g_{S_{ij}}S_{ij}$, and by repeating the procedure for the second country's entrepreneurs, as

$$(126) \qquad X_{11}(\tau)/I_{11}(\tau) = X_{21}(\tau)/I_{21}(\tau)$$

$$(127) \qquad X_{12}(\tau)/I_{12}(\tau) = X_{22}(\tau)/I_{22}(\tau)$$

for any τ. The reader will see that since $0<\beta_j<1$, the second-order condition $\partial^2\zeta_1(\tau)/\partial[S_{11}(\tau)]^2<0$ is indeed satisfied.

3. *Sector Revenues*

Divide each of the sector-output equilibrium conditions (70) through (73) by

[5] Contrary to what was assumed by Hamada (p. 66) then, the physical marginal productivities of capital in parent firm and foreign subsidiary need not be equal: In equilibrium, a rapidly growing physical capital stock will have a higher physical marginal productivity than a more slowly growing one belonging to the same owner. Does this represent underinvestment in the rapidly growing stock? It does not. Section III showed that when measured in Country 1's monetary unit, all four sector revenues $P_{ij}X_{ij}$—and hence all four amounts of money profits $\beta_jP_{ij}X_{ij}$—will grow at the same proportionate rate. Thus the rapidly growing stock will produce future amounts of money profits growing no more rapidly than those produced by the slowly growing stock.

output X_{ij}. Express the difference between the conditions for $ij=11$, 21, use (126) and find that $(C_{111}+C_{211})/X_{11}=(C_{121}+C_{221})/X_{21}$. Next express the difference between the conditions for $ij=12$, 22, use (127) and find that $(C_{112}+C_{212})/X_{12}=(C_{122}+C_{222})/X_{22}$.

Multiply each of the sector-output equilibrium conditions (70) through (73) by its price P_{ij} expressed in the same country's monetary unit for all four of them. Add the conditions for $ij=11$, 21, using (48) and (123). Next add the conditions for $ij=12$, 22, using (49) and (124). Use results above, insert (62) through (69), and find

$$P_{11}X_{11} = \frac{(\pi_{111}Y_1+\pi_{211}EY_2)R_1}{(\pi_{111}+\pi_{121})Y_1+(\pi_{211}+\pi_{221})EY_2} \tag{128}$$

$$P_{12}X_{12} = \frac{(\pi_{112}Y_1+\pi_{212}EY_2)ER_2}{(\pi_{112}+\pi_{122})Y_1+(\pi_{212}+\pi_{222})EY_2} \tag{129}$$

$$EP_{21}X_{21} = \frac{(\pi_{121}Y_1+\pi_{221}EY_2)R_1}{(\pi_{111}+\pi_{121})Y_1+(\pi_{211}+\pi_{221})EY_2} \tag{130}$$

$$EP_{22}X_{22} = \frac{(\pi_{122}Y_1+\pi_{222}EY_2)ER_2}{(\pi_{112}+\pi_{122})Y_1+(\pi_{212}+\pi_{222})EY_2} \tag{131}$$

Once again use (70) through (73), (48) and (49), (123) and (124), and (62) through (69), but this time insert (50), (51), (60), and (61) and find

$$R_1 = q_1W_1 + q_2EW_2 \tag{132}$$

$$ER_2 = q_3W_1 + q_4EW_2 \tag{133}$$

where q_i are groupings of parameters listed in Section 4 below.

Use (40) through (43) together with (56) through (59) to express the national wage bills as

$$W_1 = \alpha_1P_{11}X_{11} + \alpha_2P_{12}X_{12} \tag{134}$$

$$W_2 = \alpha_1P_{21}X_{21} + \alpha_2P_{22}X_{22} \tag{135}$$

4. *Solving for the Exchange Rate*

At long last we are now able to reduce our non-linear system of 74 equations in 74 variables to one cubic equation in one variable. To simplify that equation define

$$x = EW_2/W_1 \tag{136}$$

We may now insert (60), (61), (132), and (133) into (128) and (129), insert the results into (134), use (136), and get the cubic equation in our new unknown x:

$$\begin{aligned} &[m_1 + m_5 + (m_2 + m_6)x] \\ &\cdot[m_3 + m_7 + (m_4 + m_8)x] \\ &\quad = \alpha_1(q_1 + q_2x)(m_1 + m_2x) \\ &\qquad \cdot[m_3 + m_7 + (m_4 + m_8)x] \\ &\qquad + \alpha_2(q_3 + q_4x)(m_3 + m_4x) \\ &\qquad \cdot[m_1 + m_5 + (m_2 + m_6)x] \end{aligned} \tag{137}$$

If instead we had inserted (60), (61), (132) and (133) into (130) and (131), had inserted the results into (135) and used (136), we would have arrived at another cubic equation in our new unknown x:

$$\begin{aligned} &x[m_1 + m_5 + (m_2 + m_6)x] \\ &\cdot[m_3 + m_7 + (m_4 + m_8)x] \\ &\quad = \alpha_1(q_1 + q_2x)(m_5 + m_6x) \\ &\qquad \cdot[m_3 + m_7 + (m_4 + m_8)x] \\ &\qquad + \alpha_2(q_3 + q_4x)(m_7 + m_8x) \\ &\qquad \cdot[m_1 + m_5 + (m_2 + m_6)x] \end{aligned} \tag{138}$$

where

$$\begin{aligned} m_1 &= \pi_{111}(1 + \beta_1q_1) + \pi_{211}\beta_2q_3 \\ m_2 &= \pi_{211}(1 + \beta_2q_4) + \pi_{111}\beta_1q_2 \\ m_3 &= \pi_{112}(1 + \beta_1q_1) + \pi_{212}\beta_2q_3 \\ m_4 &= \pi_{212}(1 + \beta_2q_4) + \pi_{112}\beta_1q_2 \\ m_5 &= \pi_{121}(1 + \beta_1q_1) + \pi_{221}\beta_2q_3 \\ m_6 &= \pi_{221}(1 + \beta_2q_4) + \pi_{121}\beta_1q_2 \\ m_7 &= \pi_{122}(1 + \beta_1q_1) + \pi_{222}\beta_2q_3 \\ m_8 &= \pi_{222}(1 + \beta_2q_4) + \pi_{122}\beta_1q_2 \\ q_1 &= \{\alpha_2[1 - (\pi_{112} + \pi_{122})] \\ &\quad + \beta_2(\pi_{211} + \pi_{221})\}/D \\ q_2 &= (\pi_{211} + \pi_{221})/D \\ q_3 &= (\pi_{112} + \pi_{122})/D \end{aligned}$$

$$(139) \qquad \kappa_{11} = \frac{\beta_1 g s_{11}}{1 - [(m_1 + m_5)W_1 + (m_2 + m_6)EW_2]/(q_1 W_1 + q_2 EW_2)}$$

$$(140) \qquad \kappa_{12} = \frac{\beta_2 g s_{12}}{1 - [(m_3 + m_7)W_1 + (m_4 + m_8)EW_2]/(q_3 W_1 + q_4 EW_2)}$$

$$(141) \qquad \kappa_{21} = \frac{\beta_1 g s_{21}}{1 - [(m_1 + m_5)W_1 + (m_2 + m_6)EW_2]/(q_1 W_1 + q_2 EW_2)}$$

$$(142) \qquad \kappa_{22} = \frac{\beta_2 g s_{22}}{1 - [(m_3 + m_7)W_1 + (m_4 + m_8)EW_2]/(q_3 W_1 + q_4 EW_2)}$$

$$q_4 = \{\alpha_1[1 - (\pi_{211} + \pi_{221})] + \beta_1(\pi_{112} + \pi_{122})\}/D$$

$$D = \alpha_1\alpha_2 + \alpha_2\beta_1(\pi_{112} + \pi_{122}) + \alpha_1\beta_2(\pi_{211} + \pi_{221})$$

At this point, one is indeed tempted to assume (106) through (109) in Section IV to hold, for then $m_1 = m_5$, $m_2 = m_6$, $m_3 = m_7$, and $m_4 = m_8$. One might then divide (138) by (137) and get the linear equation $x = 1$ and our solution (110) for the exchange rate.

5. *Solving for the Marginal Productivities of Capital*

Use the definitions (44) through (47) and (1) through (35) to express the marginal productivities of capital in terms of the ratio X_{ij}/I_{ij}. Then use (60) through (73) and (128) through (133) to express that ratio. Without using our special case (106) through (109) we may then find the four marginal productivities of capital as shown in equations (139)–(142), above.

In the special case (106) through (109), our results (139) through (142) collapse into (119) through (122) in Section IV.

6. *Solving for Net National Capital Export*

First use the definitions (44) through (47) and (1) through (35) to express investment in terms of output and (139) through (142):

$$(143) \qquad P_{12}I_{12} = P_{12}X_{12}\beta_2 g s_{12}/\kappa_{12}$$

$$(144) \qquad EP_{21}I_{21} = EP_{21}X_{21}\beta_1 g s_{21}/\kappa_2.$$

Second, recall that equations (123) and (124) defined the savings of Country 1 and Country 2, respectively. The combined investments undertaken within those countries are, respectively,

$$(145) \qquad P_{11}I_{11} + P_{12}I_{12}$$

$$(146) \qquad EP_{21}I_{21} + EP_{22}I_{22}$$

Now deduct (145) from (123) and find Country 1's excess of saving over investment within it or, which is the same thing, Country 1's net national capital export

$$(147) \qquad EP_{21}I_{21} - P_{12}I_{12}$$

Had we deducted (146) from (124) we would have found Country 2's net national capital export equalling, of course, (147) with opposite sign.

REFERENCES

R. J. Ball and K. Mavwah, "The U.S. Demand for Imports, 1948-1958," *Rev. Econ. Statist.*, Nov. 1962, *44*, 395-401.

G. H. Borts, "A Theory of Long-Run International Capital Movements," *J. Polit. Econ.*, Aug. 1964, *72*, 341-59.

——— and J. L. Stein, *Economic Growth in a Free Market*. New York 1964.

F. D. Graham, "The Theory of International Values Re-Examined," *Quart. J. Econ.*, Nov. 1923, *38*, 54-86.

K. Hamada, "Economic Growth and Long-Term International Capital Movements," *Yale Econ. Essays*, spring 1966, *6*, 49-96.

A. C. Harberger, "Some Evidence on the International Price Mechanism," *J. Polit. Econ.*, Dec. 1957, *65*, 506-22.

H. S. Houthakker and S. P. Magee, "Income and Price Elasticities in World Trade," *Rev. Econ. Statist.*, May 1969, *51*, 111-25.

M. Kreinin, "Price Elasticities in International Trade," *Rev. Econ. Statist.*, Nov. 1967, *49*, 510-16.

H. Lund, *Svenska företags investeringar i utlandet.* Stockholm 1967.

E. L. Nelson and F. Cutler, "The International Investment Position of the United States in 1967," *Surv. Curr. Bus.*, Oct. 1968, *48*, 19-32.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 65-94.

D. S. Swamy, "Statistical Evidence of Balanced and Unbalanced Growth," *Rev. Econ. Statist.*, Aug. 1967, *49*, 288-303.

Deutsche Bundesbank, "Kapitalverkehr mit dem Ausland," *Monatsberichte der Deutschen Bundesbank*, Apr. 1969, *21*, 64.

Land and Economic Growth

By DONALD A. NICHOLS*

This paper incorporates land in a neoclassical model of economic growth. Saving and investment functions are modified due to the existence of land and these modifications yield some interesting results concerning the rate of capital accumulation: First, the maximum consumption path is unattainable; and second, the rate of capital accumulation depends negatively on both the equilibrium rate of growth and the relative share of land in national income. The latter result is a quantification of an effect claimed by many observers to characterize certain underdeveloped countries; namely, that saving motives are satisfied by land holdings (and the increase in real land prices) rather than by capital accumulation.

Equilibrium in neoclassical growth models with two assets was first examined by James Tobin (1965). Most of the subsequent work on two asset models has continued to use Tobin's assumption that the asset other than capital is government debt and that its importance for equilibrium growth obtains solely from its role as a store of value.[1] Don Patinkin and David Levhari recently extended this analysis by considering money to be a productive asset. Their results differ from those of this paper primarily because of the nature of money; namely, that its yield of real transaction services varies with its price. This is not true of any other asset.

The model is developed in Section I and the existence of an equilibrium is demonstrated. The result concerning the maximum consumption path is then exhibited. Section II explores the quantitative effect of rents and growth rates on capital accumulation, while Section III consists of an analysis of the stability of the model under alternative assumptions concerning the formation of expectations about the price of land. In Section IV, alternative assumptions about technical progress are discussed, and Section V contains a summary of the results and some further implications of the model.

I. *The Model*

There is one kind of output in this model which can be consumed or invested. The production function for this output (Q) is twice differentiable, homogeneous of degree one with land (L), labor (N), and capital (K) as productive factors. Technological change exists and is assumed to be land and labor augmenting and of such a character that the geometric growth rates of augmented land (L^*) and augmented labor (N^*) are equal.[2] Call this growth rate g. Thus: $L^* = Le^{gt}$, $N^* = N^*(0)e^{gt}$.

The production function can be written as a function of K, N^* and L^*.

$$Q = F(K, N^*, L^*); \qquad F_{K,N^*,L^*} > 0; \qquad F_{KK,N^*N^*,L^*L^*} < 0 \tag{1}$$

* The author is assistant professor of economics at the University of Wisconsin. He wishes to thank D. D. Hester, J. S. Chipman, J. E. Floyd, C. W. Reynolds, and D. Keesing for their comments. He gratefully acknowledges the financial support of the National Science Foundation and of the Research Committee of the Graduate School of the University of Wisconsin.

[1] See H. G. Johnson, M. Sidrauski (1967a), P. A. Diamond, E. A. Thompson, and J. E. Stiglitz, Sidrauski and K. Shell.

[2] This assumption can be justified by reference to the literature concerning the endogenous bias of technical change developed by C. Kennedy, E. S. Phelps, E. M. Drandakis, and J. S. Chipman. According to this literature, profit maximization leads firms to direct innovative activity toward economizing on the most slowly growing factors. In equilibrium, all augmented factors

Wealth (W), measured in terms of goods, equals the capital stock plus the value of land. P is the price of land in terms of goods.

$$(2) \qquad W = PL + K$$

Saving (S) is defined as the increase in wealth. Following the previous literature, saving is assumed to be a linear function of disposable income, where disposable income includes capital gains on land.[8]

$$(3) \; S \equiv DW \equiv DK + (DP)L = S[Q + (DP)L]$$

The D operator represents the time derivative.

Equation (3) has an important influence on our results. It implies that the motive for saving is to increase wealth. This motive can be satisfied either by capital accumulation or by increases in the value of land. Thus the rate of capital accumulation and the rate of increase of land price are simultaneously determined in the model with (3) restricting their weighted sum to be equal to a certain percentage of output. Our result concerning the effect of land on capital accumulation depends importantly on the form of this equation. It should be clear that large quantities of land—and therefore large capital gains from any price change—can lead to low levels of capital accumulation.

The saving equation can be viewed as determining the desired level of wealth as a function of the level of output. Because there are two assets, an additional equation is needed to portray the portfolio decisions of wealth holders. This could be either an investment equation which showed the demand for capital, an equation determining the demand for land or an equation to explain the price of land. The portfolio equation we have chosen determines the price of land by requiring that it be equal to the present value of the future returns to land. As capital is the only asset other than land, the rate of return on capital is used as a discount rate.

$$(4) \quad P(\mathrm{t}) = \int_{\mathrm{t}}^{\infty} F_L(v) \exp \left[- \int_{\mathrm{t}}^{v} F_K(u) du \right] dv$$

These four equations; the production function, the saving function, the investment function, and the wealth identity are functions of four variables: Q, K, W, and P. They can be reduced to two differential equations in two unknowns. We now examine the solution to these two equations leaving the discussion of stability until Section III.

Balanced Growth

Define $k = K/N^*$ to be the ratio of capital to effective labor inputs. Define $l = PL/N^*$ to be the ratio of land to labor inputs. Note that from (4), the price of land is determined by the value of land as a productive factor, thus P should grow with land-augmenting technological change.

A balanced growth path is defined as a constancy in k and l. Thus $Dk/k = Dl/l = 0$ will characterize the balanced growth equilibrium.

Let us first examine the conditions for $Dl/l = 0$.

$$(5) \quad \begin{aligned} Dl/l &= DP/P + DL/L - DN^*/N^* = 0 \\ &= DP/P + 0 - g = 0 \end{aligned}$$

Differentiating (4) with respect to time and dividing by $P(\mathrm{t})$ gives $DP(\mathrm{t})/P(\mathrm{t})$.

$$(6) \qquad \frac{DP(\mathrm{t})}{P(\mathrm{t})} = - \frac{F_L(\mathrm{t})}{P(\mathrm{t})} + F_K(\mathrm{t})$$

grow at the same rate. These results require that the elasticity of substitution be bounded below one.

[8] Tobin (1965) assumed saving to be a linear function of output plus new issues of government debt, while Patinkin and Levhari and Shell, Stiglitz and Sidrauski include capital gains on existing debt as a part of disposable income.

Equation (6) tells us that the marginal efficiency of capital must equal the marginal efficiency of land plus the rate of increase in land price. That is, the rates of return on the two assets must be equal and part of the rate of return on land is its rate of price appreciation. Substituting (6) into (5), we get (7)

$$-\frac{F_L}{P} + F_K - g = 0 \tag{7}$$

Noting that $L^*(t)$ is some multiple of $N^*(t)$, it is evident that the marginal product of capital can be expressed as a function of the capital-augmented labor ratio only. Letting $N^*(t)=\beta L^*(t)$, remembering that $L^*(t)=Le^{gt}$, that $F_L = F_{L^*}e^{gt}$, and that F_{L^*} depends on the capital-augmented labor ratio, we can rewrite (7) as (8), an equilibrium relationship between k and l.

$$\begin{aligned} &-\frac{F_{L^*}(k)}{l\beta} + F_K(1/k) - g = 0 \\ &\therefore\ l = \frac{F_{L^*}(k)}{(F_K(1/k) - g)\beta} \end{aligned} \tag{8}$$

Clearly, $\partial l/\partial k > 0$. Thus (8) tells us that l is a monotonic function of k. Moreover, l goes to infinity for some finite value of k, namely for that value of k which sets the rate of return on capital equal to the growth rate.

Intuitively, (8) results from the equation which says that the rates of return on the two assets must be equal. The higher the capital intensity (k), the lower will be the rate of return on capital. If the rate of return on land is also to be lower, the price of land must be higher since the increased capital intensity increases the marginal product of land. But part of the return to land is the increase in the price of land. As this occurs at geometric rate g, the price of land will tend to infinity as the rate of return on capital approaches g. It becomes clear below that it is this factor which prevents the maximum consumption path from being attained.

Saving behavior yields the other long-run relationship between k and l.

$$\begin{aligned} Dk/k &= DK/K - DN^*/N^* \\ &= DK/K - g = 0 \end{aligned} \tag{9}$$

From (3) we can derive DK/K.

$$DK + (DP)L = s[Q + (DP)L] \tag{3}$$

$$\begin{aligned} DK/K &= sQ/K + (s-1)(DP)L/K \\ &= sF(1/k) + (s-1)\frac{DP}{P}\cdot l/k \end{aligned} \tag{10}$$

From (5) we know that $DP/P = g$ along a balanced growth path, therefore to evaluate (10) at the equilibrium, we can set $DP/P = g$. Thus (11) becomes the relationship that defines a constancy in the capital-labor ratio.

$$\begin{aligned} &Dk/k = sF(1/k) + (s-1)gl/k - g = 0 \\ &\therefore\ l = \frac{-sF(1/k)k}{(s-1)g} + \frac{k}{s-1} \end{aligned} \tag{11}$$

Note that for $l=0$, this yields the Solow-Swan result: $sF(1/k)=g$.[4] Differentiating with respect to k, we find that $\partial l/\partial k$ is positive for small k, negative for large k.[5]

$$\begin{aligned} \partial l/\partial k &= \frac{-sF'(k) + g}{(s-1)g} < 0 \\ &\textit{for } F'(k) < g/s \end{aligned} \tag{12}$$

Graphing (11) and (8) we see the existence of an equilibrium.[6] Before discussing stability, we will note the characteristics of the equilibrium path.

[4] Actually, there are two roots for $l=0$, the other being $k=0$.

[5] This is easily seen by noting that $F'(k) < g/s$ makes (12) negative. But $F'(k) < F(1/k)$ due to diminishing returns and $F(1/k) = g/s$ for $l=0$. For smaller k, $F'(k)$ is larger.

[6] Equilibrium is guaranteed by the fact that $\partial l/\partial k = \infty$ when $k=0$ in (12), while $\partial l/\partial k = 0$ when $k=0$ in (8). This results from our assumptions about the production function.

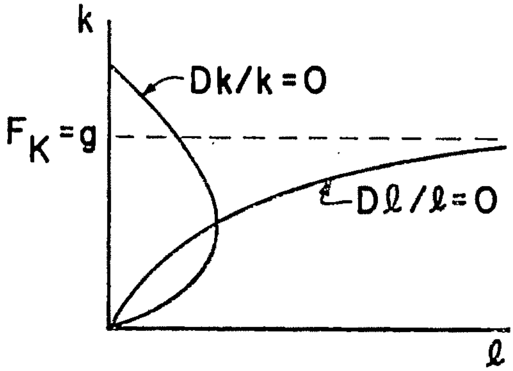


FIGURE 1

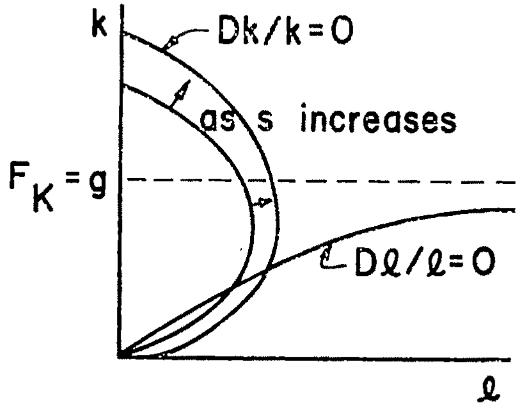


FIGURE 2

The Maximum Consumption Path

Equation (8) tells us that F_K must always exceed g if l is to be finite. From Phelps we know that this makes the maximum consumption path (characterized by $F_K=g$) unattainable. We also know that this prevents an inefficient accumulation of too much capital. The introduction of land as an alternative store of purchasing power has made over-saving unnecessary even if people wish to have very large levels of wealth.

Desires for large levels of wealth will be expressed in high saving rates. Figure 2 shows how increases in saving rates increase both k and l. Wealth increases for two reasons: more capital is accumulated, and the price of land increases. The increase in the price of land will be large enough to satisfy any saving motive without accumulating capital to the point where $F_K=g$.

The reason for this response in k and l is that when more capital is accumulated, its rate of return falls. Portfolio balance requires land to have the same rate of return as capital. But with the supply of land fixed, the marginal product of land cannot vary except in response to changes in technology and factor ratios (such as k). The increase in k caused by higher saving rates increases the marginal product of land. If the rate of return on land is to fall with that on capital, the price of land must increase. For very high saving rates, the rate of return on capital may fall to levels near the growth rate. But the price of land must be very high at this point to preserve equilibrium.

At no finite saving rate can the rate of return on capital exceed or equal the growth rate. An infinite level of l is necessary if the growth rate equals the rate of return.

A similar result was anticipated by Keynes.

> It may be that in certain historic environments the possession of land has been characterized by a high liquidity-premium in the minds of owners of wealth; and since land resembles money in that its elasticities of production and substitution may be very low, it is conceivable that there have been occasions in history in which the desire to hold land has played the same rôle in keeping up the rate of interest at too high a level which money has played in recent times. [p. 241]

The quote from Keynes points up the fact that it is the portfolio balance equation which guarantees our result concerning the maximum consumption path. A more sophisticated function which allowed the rates of return on capital and land to differ could lead to a different result if the

rate of return on capital required by wealth holders was less than that required on land.

II. *Land and Saving Rates*

In this section, we explore the quantitative implications of the model. We had intended to perform a cross-country comparison of saving rates as affected by the inclusion of land. But, necessary for such calculations were data on aggregate returns to land. Most observed rents include payments for buildings as well as land; therefore, most statistical sources report a composite rent. There is, then, no simple way to use this theory to explain national income accounts data.

We can, however, get an idea of the magnitudes involved by looking at hypothetical numbers. That is, we can see how the growth rate, the saving rate, and the factor shares relate to each other quantitatively. In this section, I will show how different levels of rents may affect observed saving behavior if long-run equilibrium is thought to characterize an economy.

Dividing equation (3) by Q, substituting gP for DP, $F_L/(F_K-g)$ for P, and letting $R=F_LL/Q$, the relative share of land, we get (13) which shows us how saving rates as measured in the national income accounts are functions of the relative share of land. National income accounts saving includes only capital accumulation. Capital gains are not counted as a part of income or saving. The saving function outlined above includes capital gains in income and saving. The saving rate above is denoted s; the national income accounts saving is DK/Q. They are related by (13).

$$\text{(13)} \quad DK/Q = s - (1-s)\frac{g}{F_K - g}R$$

To see what (13) implies for some representative values of the parameters, let us assume the following: $F_K \doteq .2$, $g=.05$, $s=.10$. If we substitute these into (13), we get a function that relates the rate of capital accumulation to the relative share of rents as shown in (13a).

$$\text{(13a)} \quad DK/Q = .1 - .9\frac{.05}{.15}R = .1 - .3R$$

For $R=.1$, DK/Q will equal .07. For $R=.2$, DK/Q will be .04 while an R of .3 implies a DK/Q of .01. That is, if rents are a large part of GNP, increases in land value caused by the growth of capital and labor will satisfy a large part of the desire for saving. Therefore, even if all countries had the same saving motives, one would expect that those countries in which rents are a large part of GNP would have low levels of capital accumulation.

This result is not too surprising given the assumptions concerning technical change. Land-augmenting technical progress is effectively the creation of new land. The larger the share of land in national income, the greater the value of the new land created. This new land is a part of output and saving in the model developed above.

The lower the profit rate and the higher the growth rate, the more striking this result becomes. For example, if the profit rate, F_K, is 10 percent while the growth rate remains at 5 percent, we then have $DK/Q=.1-.9R$; so that rents equal to 10 percent of national income reduce the observed rate of capital accumulation to 1 percent of income.

This analysis implies that saving in the form of land can exist and can be very large relative to saving in the form of capital accumulation. One policy implication of this analysis is that attempts to increase rates of capital accumulation in countries with large quantities of rents are more likely to be successful if rents are taxed than otherwise. Taxing rents should lower the price of land and therefore the amount of capital gains on land which result from economic growth. To

satisfy the same saving motives as before the tax was imposed will require an increase in the rate of capital accumulation.

III. *Stability*

We will now examine the local stability of the equilibrium and note the importance of assumptions about speculative behavior for that stability.

Local stability of the equilibrium exists when the two requirements of (14) are satisfied.

$$(14)\quad \begin{aligned} &\frac{\partial(Dl)}{\partial l}+\frac{\partial(Dk)}{\partial k}<0\\ &\frac{\partial(Dl)}{\partial l}\cdot\frac{\partial(Dk)}{\partial k}-\frac{\partial(Dl)}{\partial k}\cdot\frac{\partial(Dk)}{\partial l}>0\end{aligned}$$

Substituting DP/P back into (11), we can use (8) and (11) to derive the derivatives represented by (15), from which it is clear that the second of the stability requirements does not hold in the neighborhood of the equilibrium.

$$(15)\quad \begin{aligned} \partial(Dl)/\partial l &= F_k(1/k)-g && >0\\ \partial(Dk)/\partial k &= sF_K(1/k)-g \\ &\quad +(s-1)\partial(Dl)/\partial k && \gtrless 0\\ \partial(Dl)/\partial k &= \frac{-F_{L*k}}{\beta}+F_{Kk}l && <0\\ \partial(Dk)/\partial l &= (s-1)\partial(Dl)/\partial l && <0\end{aligned}$$

Thus if we use (8) and (11) to represent the dynamic behavior of the model off the equilibrium path as well as on it, the model is locally unstable.[7] But (8) is derived from the portfolio balance equation which requires that the rates of return on the assets be equal. A part of the rate of return on land is its increase in price. Not surprisingly, this gives the result that increases in the price of land imply that a higher price of land is necessary for the rates of return on the two assets to be equal. Thus if the price of land ever deviates from the level which makes $Dl/l=0$, it continues to move further from this value. Equation (8) therefore incorporates speculative behavior of a form which implies that all capital gains on land must be capitalized in the price of land. It is as if asset holders expected any rate of change of price to continue forever.

Such speculative behavior is possible although not frequently observed. Rarely do we observe the price of any asset tending toward infinity with the demand for the asset increasing as a result of the potential capital gains. The historical examples available—such as the tulip bulb boom or the South Sea bubble—are rare, albeit possible, and all have eventually ended.

An alternative model of speculation is possible. One need not assume that investors capitalize short-term capital gains. In this alternative model, equilibrium price is determined by the present value of the real returns. Price increases above that level (and thus a temporary rate of return above the equilibrium level) are signals to sell the asset, not to buy more of it.

This alternative model is merely a statement about expectations of investors; namely, that they expect the price of an asset to be ultimately determined by its real returns. We now use this model to examine the stability of the equilibrium implied by (8) and (11).

Along any balanced growth path, (4) tells us that the price of land equals the discounted real returns where these re-

[7] It should be noted that the conclusion by Patinkin and Levhari (p. 745) that stability is possible in their model results from the fact that when examining the stability of the single equation $Dk/k=g(k)$, they require that the other equation be in equilibrium although they never examine its stability. That is, they plug in $Dm/m=0$ (analagous to our Dl/l equation) into the Dk/k equation and then ignore the Dm/m equation for stability purposes. This implicitly assumes that the stability of the system is unaffected by the nature of the speculative behavior of asset holders and that this behavior is such as to guarantee that $Dm/m=0$ is always satisfied.

turns are expected to grow at rate g. This equilibrium price is noted in (16).

$$P(\mathrm{t}) = \frac{F_L(\mathrm{t})}{F_k - g} \tag{16}$$

Assuming that investors will buy land if its price is less than that exhibited by (16) and sell it if its price exceeds (16), we get (17) as an alternative explanation of short-run price movements.

$$DP/P = \gamma\left(\frac{F_L}{F_k - g} - P\right) + g \tag{17}$$

Equation (17) can be thought of in light of the correspondence principle as a corresponding dynamic model. We could think of the behavior it describes as occurring very rapidly relative to the dynamic behavior of the equilibrium model. It is a corresponding disequilibrium model that accounts for short-run speculative price movements.

Stability analysis using (17) can be thought of as answering the following question. Can the balanced growth path be stable if speculators use expected real returns to determine the price of land? That is, we are now asking whether there is a tendency to a long-run equilibrium capital-labor ratio as long as the short-run speculative behavior that determines the relative prices of assets is stabilizing. If we think of (17) as describing instantaneous behavior, (γ is very large relative to the other dynamic parameters), then our answer to this question is clearly yes. Inserting (17) into (5) and differentiating with respect to k and l, we get (18).

$$\begin{aligned} &\partial(Dl)/\partial l = -\gamma && <0 \\ &\partial(Dl)/\partial k \\ &\quad = \gamma\left(\frac{F_{L^*k}}{\beta(F_k - g)} - \frac{F_{kk}F_{L^*}}{\beta(F_k - g)^2}\right) && >0 \end{aligned} \tag{18}$$

Substituting (18) into (14) along with the derivatives of Dk shown in (15), we find the solution to be unambiguously stable. Thus the model can exhibit a long-run stable equilibrium if there is short-run stabilizing behavior in the asset markets.

IV. *Disclaimers*

It should be made clear that the assumption concerning land-augmenting technological change is an important ingredient in the existence of a balanced growth solution to the model. The induced invention models of Kennedy, Phelps, Drandakis, and Chipman were mentioned as justification for this assumption. It is still possible, of course, that land augmenting technological change of the appropriate magnitude will not be forthcoming. If, for example, no land augmenting progress occurs, our results change as follows.

With augmented factors growing at different rates the factor prices will not attain a constant value. It is well known that if the elasticity of substitution of factors (σ) in the production function is less than one, these factor price changes will be of such a character that the relative share of the most slowly growing factor will approach one; if $\sigma > 1$, its share will approach zero, while $\sigma = 1$ implies constant factor shares. Thus in the absence of the appropriate technological progress, the price of land will grow faster than or slower than the growth rate of output in a manner that depends upon the elasticity of substitution. If, by chance, the production function were Cobb-Douglas, constant factor shares will result but will not correspond to a balanced growth path because factor prices will be changing. With no labor-augmenting technical change, for example, wage rates would be falling. Capital, of course, will grow at the rate of growth of output which will be a weighted average of the rates of growth of

land and labor. Interest rates could therefore be constant. Thus even in the absence of land-augmenting technical change, a quasi-balanced growth path can result if the production function is Cobb-Douglas. It is *quasi* instead of *real* because wage rates increase not at the rate of labor-augmenting technical progress but at that rate minus the difference in the rates of growth of augmented labor and output.

A further disclaimer is that land, like capital and labor, is not homogeneous. Technical progress will tend to increase the value of some resources while decreasing that of others. Even if the simple story depicted above is a useful way to characterize the world's economy, an unequal distribution of the various components of land across the different nations may make the model a less useful characterization of national economies.

V. *Conclusion*

If there exists a productive factor which is fixed in supply—such as land—and if technological change is of the proper character, then a balanced growth path can be shown to exist given the usual neoclassical assumptions. One key property of the balanced growth path will be that the price of the fixed factor will increase at a geometric rate equal to the rate of growth of output. The resulting capital gains may be considered as income and saving by asset holders and therefore they may diminish the demand for wealth in the form of real capital. In this paper, land has been used as the example; capital gains resulting from the increase in land prices can satisfy the desires for increased wealth. Saving in the form of land can be shown to affect the level of real capital accumulation and the magnitude of the effect depends on the rate of growth and the share of land in national income.

Because the price of land must increase proportionately at the growth rate of output while the rental yield is also positive, the rate of return on land must exceed the growth rate. If asset holders require that the rates of return on all assets be equal, then the rate of return on capital must also be at least as large as the growth rate. This yields as a result the fact that the maximum consumption path is no longer attainable. Oversaving or inefficient accumulation of excess capital in the sense of Phelps is shown to be impossible.

The model was shown to be unstable if wealth holders capitalize all short-term price movements and stable if they don't. The introduction of a corresponding disequilibrium model to account for short-run price fluctuations was shown to have a substantial effect on the stability of the system. It is likely that the introduction of such speculative behavior in other models which have been found to be unstable will affect them in the same fashion. Thus in this model, we were able to exhibit the existence of a stable long-run balanced growth path once we ruled out the possibility of speculative price movements toward zero or infinity.

A further implication of the above model should be noted. A change in the rate of growth of output leads to a once and for all capital gain for land holders. This is because the returns to land are capitalized at a rate equal to the marginal product of capital minus the growth rate, (or alternatively, because the future returns to land can be expected to be larger with a high growth rate than with a low one). This may explain the capital gains in real estate that exist in cities whose growth rates increase suddenly, or even in farm land as it becomes residential real estate if the output of the city core is expanding more rapidly than the value of agricultural output.

REFERENCES

J. S. Chipman, "Induced Technical Change and Patterns of International Trade," presented at the Conference on Technology and Competition in International Trade, New York, Oct. 11–12, 1968.

P. A. Diamond, "National Debt in a Neoclassical Growth Model," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1126–50.

H. G. Johnson, "The Neo-Classical One-Sector Growth Model: A Geometrical Exposition and Extension to a Monetary Economy," *Economica*, Aug. 1966, *33*, 265–87.

C. Kennedy, "Induced Bias in Innovation and the Theory of Distribution," *Econ. J.*, Sept. 1964, *74*, 541–47.

J. M. Keynes, *The General Theory of Employment, Interest, and Money*. New York 1936.

D. Levhari and D. Patinkin, "The Role of Money in a Simple Growth Model," *Amer. Econ. Rev.*, Sept. 1968, *58*, 713–53.

E. S. Phelps, "Second Essay on the Golden Rule," *Amer. Econ. Rev.*, Sept. 1965, *55*, 793–814.

E. S. Phelps and E. M. Drandakis, "A Model of Induced Invention, Growth and Distribution," *Econ. J.* Dec. 1966, *76*, 823–29.

K. Shell, M. Sidrauski and J. E. Stiglitz, "Capital Gains, Income and Saving," *Rev. Econ. Stud.*, Jan. 1969, *36*, 15–20.

M. Sidrauski, (a.) "Inflation and Economic Growth," *J. Polit. Econ.*, Dec. 1967, *75*, 796–810.

———, (b.) "Rational Choice and Patterns of Growth in a Monetary Economy," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 534–44.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 65–94.

E. A. Thompson, "Debt Investments in Both Macroeconomic and Capital Theory," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1196–1210.

J. Tobin, "Money, Capital and Other Stores of Value," *Amer. Econ. Rev. Proc.*, May 1961, *51*, 26–37.

———, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–84.

Cross-Section and Time-Series Tests of the Permanent-Income Hypothesis

By JULIAN L. SIMON AND DENNIS J. AIGNER*

Soon after Keynes' General Theory appeared, cross-sectional budget studies were adduced to support Keynes' "psychological law" of decreasing average consumption with increasing income. When Milton Friedman (hereafter **F**) and Franco Modigliani et al. (hereafter **M**) wished to argue the contrary, that consumption is proportional to income independent of the income level, they found it necessary to explain away the cross-sectional budget-study evidence.[1]

First **F** and **M** showed how it was *possible* for a cross-section regression of measured consumption on measured income to *show* a lower marginal propensity to consume (*MPC*) than average propensity to consume (*APC*) even if the "true" *MPC* is equal to *APC*. The explanation lies in the effects of errors in variables on least squares regression, in this case the "errors" being the year-to-year variations in family income. If consumption depends on income over more than one year, a one year cross-section must show a lower apparent *MPC* than the true *MPC*. But neither **F**'s nor **M**'s explanations were clear or neat. Hence the first task of this paper (Section I) is to show the logic of a simple way by which *APC* can be reconciled to the short-run *MPC*, as shown by cross-section plots of current consumption on current income. This is the task to which both **F** and **M** devoted much effort and space. **F** invented the concepts "transitory income" and "permanent income" to unravel the paradox, but the following development is much simpler than the previous complicated constructions, and renders those concepts unnecessary. Edmund Malinvaud (pp. 126–29) preceded us in this direction, building on the work of Yehuda Grunfeld.

F then went beyond his speculation and made various manipulations of cross-section budget studies which suggested that the cross-section data are indeed compatible with the errors-in-variables explanation. But all that **F**'s test could show was a *qualitative* result—for example, that the one year cross-sections err in the direction of the **F** and **M** hypotheses of a "horizon" longer than one year and of proportionality. The tests certainly do not show that all or even most of the discrepancy between *APC* and *MPC* may be explained this way.[2] But one certainly wishes to have some quantitative estimate of the extent to which the removal of the statistical distortion in one-period cross-sections brings together *APC* and *MPC*. The logic

* University of Illinois and University of Wisconsin. The authors are grateful to Milton Friedman, Martin Ringo, and a referee for helpful comments.

[1] The ideas of Friedman's that are referred to here are mostly contained in his book, an article which appeared in 1963, and a comment or two. The ideas of the group identified with Modigliani were published earliest by Modigliani and Brumberg. Subsequent empirical work was done by Modigliani and Ando (1957, 1960), and further theoretical discussion and empirical work appeared in Modigliani's 1966 article.

[2] Friedman does provide "tests" on at least two occasions (1957, p. 97ff; pp. 196–97) which could be manipulated to yield a quantitative result, although he does not pursue the possibility.

outlined in the first section of this paper enables us to make a quantitative estimate of the relationship between APC and MPC. However, this estimation technique requires input estimates of the relative importance of income in the various past periods upon (expectations of future income and hence upon) current consumption. **F**'s own lag-structure coefficients are used. Thus the final estimate is a joint test of those coefficients *together with* the proportionality hypothesis, rather than of either alone. This work is found in the second section of the paper.

The third section of the paper carries out the same sort of analysis, though more briefly, for time-series regressions without lagged terms that have been adduced as evidence about MPC. This offers an explanation for the result that long time-series show higher MPC's than short time-series.

The fourth section then asks *which* lag coefficients would have to be the true ones if lagged effects (errors in variables) are to explain *all* the discrepancy between MPC's and APC's observed in cross-sectional budget-studies and in time-series. The reader may then judge whether such coefficients are plausible. If they are not, then the argument is against the proportionality hypothesis (assuming the basic consumption-function model makes sense).

I. *The Basic Logic*

The most basic formulation of the consumption function is that of **M**, which we will interpret as:[3]

$$C_{i,\mathrm{T}} = \frac{1}{L}\,[A_{i,\mathrm{T}} + Y_{i,\mathrm{T}} + Y'_{i,\mathrm{T}+1} + Y'_{i,\mathrm{T}+2} \cdots Y'_{i,\mathrm{T}+N}] \tag{1}$$

The symbols used above and in the rest of the paper are as follows:[4]

$Y_{i,\mathrm{T}}$ = measured income of family i in year T
$Y'_{i,\mathrm{T}}$ = expected income of family i in year T
$Y_{i,p}$ = permanent income of family i in year T
$C_{i,p}$ = permanent consumption of family i in year T
$C_{i,\mathrm{T}}$ = measured consumption of family i in year T[5]
$A_{i,\mathrm{T}}$ = assets of family i in year T
L = number of years of life expected, T and thereafter
N = number of years of working expected, T and thereafter ($L \geq N$)

One obtains a reconciliation of the work of **F** and **M** (described at length in Simon's paper) by converting **F**'s basic empirical micro-model (1963, p. 21) which is:

$$C_{i,p} = kY_{i,p} = b_0 Y_{i,\mathrm{T}} + b_1 Y_{i,\mathrm{T}-1} + b_2 Y_{i,\mathrm{T}-2} \cdots + b_J Y_{i,\mathrm{T}-J} \tag{2}$$

into an estimator of expected income.[6] If income is a stationary stochastic process, then $Y'_{i,\mathrm{T}} = Y'_{i,\mathrm{T}+N}$, which leads to

$$Y_{i,p} = Y'_{i,\mathrm{T}} = a_0 Y_{i,\mathrm{T}} + a_1 Y_{i,\mathrm{T}-1} + a_2 Y_{i,\mathrm{T}-2} \cdots + a_J Y_{i,\mathrm{T}-J} \tag{3}$$

where $a_0 = b_0/k$, $a_1 = b_1/k$, etc. By substitution of (3), **M**'s model (1) may be rewritten as

$$C_T = \frac{1}{L}\,[A_\mathrm{T} + Y_\mathrm{T} + (a_0 Y_\mathrm{T} + a_1 Y_{\mathrm{T}-1} + \cdots + a_J Y_{\mathrm{T}-J}) + (a_0 Y_\mathrm{T} + a_1 Y_{\mathrm{T}-1} + \cdots + a_J Y_{\mathrm{T}-J}) \cdots + (a_0 Y_\mathrm{T} + a_1 Y_{\mathrm{T}-1} + \cdots + a_J Y_{\mathrm{T}-J})],$$

[3] The warrant for this interpretation is Modigliani and Brumberg (equation II.1, p. 397).

[4] In the text the subscript i is often dropped for notational convenience.

[5] Permanent and measured consumption may be considered equal on the average for any group of people

or, collecting coefficients on like income terms into composites $a'_1, \ldots, a'_J$:

$$C_T = \frac{1}{L} A_T + a'_0 Y_T + a'_1 Y_{T-1} + a'_2 Y_{T-2} \cdots + a'_J Y_{T-J} \tag{4}$$

In (4) the observation subscript has been dropped for convenience. A commonly used operational model is (4) without the asset term, which leads us back to (2), written in more conventional form as the regression equation

$$C_T = b_0 Y_T + b_1 Y_{T-1} + \cdots + b_J Y_{T-J} + u, \tag{5}$$

where u is "transitory consumption," a random variable assumed to have classical attributes over the population of family units. It is (5) that forms the basis of further analysis for us in this paper.

A cross-section budget study may be used to produce a least squares slope estimate, say $\hat{b}_{0,c}$, of the relation between current consumption and current income, where the subscript (c) in $\hat{b}_{0,c}$ is used to indicate that the estimate is obtained from a cross-section. Now we use an extension of a basic result developed elsewhere (see Grunfeld, and Malinvaud, p. 127). If $\hat{b}_{0,c}$ is the least squares slope coefficient from a cross-section plot of current consumption $C_{i,T}$ on current income $Y_{i,T}$, then in view of equation (5) its expected value, $E(\hat{b}_{0,c})$, is given by:

$$E(\hat{b}_{0,c}) = b_0 + \gamma_{Y_TY_{T-1}} b_1 + \gamma_{Y_TY_{T-2}} b_2 + \cdots + \gamma_{Y_TY_{T-J}} b_J, \tag{6}$$

where $\gamma_{Y_TY_{T-j}}$ is the (simple) regression coefficient obtained by regressing Y_{T-j} on Y_T. Equation (6) is the result of an analysis of specification bias in least squares regression due to the omission of lagged variables.

Generally we would expect the coefficients $\gamma_{Y_TY_{T-j}}$ to be positive and less than one in absolute value. This will be the case, for example, if income is generated as a first-order stationary autoregressive process, $Y_T = \rho Y_{T-1} + v_T$, where v_T is "white noise," and $0<\rho<1$. Even if some of these "auxiliary" regression coefficients are negative, a modification to (6) is instructive: given stationarity, $\gamma_{Y_TY_{T-j}} \cong r_{Y_TY_{T-j}}$, the simple correlation coefficient between Y_T and Y_{T-j} in the sample.[7] Then:[8]

$$E(\hat{b}_{0,c}) \cong b_0 + r_{Y_TY_{T-1}} b_1 + r_{Y_TY_{T-2}} b_2 \cdots + r_{Y_TY_{T-J}} b_J \tag{7}$$

In a society that (for convenience) has no aggregate income trend, and in which

that have the same permanent income, by F's assumptions. (See Friedman, p. 35.)

[6] The presence of a trend term in equation (2) would in no way affect the substance of our subsequent analysis, but it would complicate the expressions to follow. It is disregarded for convenience only.

[7] The "approximation" becomes an exact relation as sample size (n) is increased. For,

$$\gamma_{Y_TY_{T-J}} = r_{Y_TY_{T-J}} \left(\sum^n Y^2_{T-J} \Big/ \sum^n Y^2_T \right)^{1/2}$$

Given stationarity,

$$\frac{1}{n} \sum^n Y^2_{T-J} \quad \text{and} \quad \frac{1}{n} \sum^n Y^2_T$$

are both consistent estimators of the same population variance. Thus in the limit, as $n \to \infty$, $\gamma_{Y_TY_{T-J}} = r_{Y_TY_{T-J}}$

[8] If A is not disregarded,

$$E(\hat{b}_{0,c}) \cong r_{Y_TA_T} \left(\frac{1}{L}\right) + b_0 + r_{Y_TY_{T-1}} + r_{Y_TY_{T-2}} b_2 \cdots + r_{Y_TY_{T-J}} b_J. \tag{7a}$$

It should be noted that equation (7) is quite different from Friedman's equation 4.1 (p. 110), $b = kP_y$ [where b is Friedman's notation for our $E(\hat{b}_{o,c})$]. P_y is estimated by Friedman on the basis of his preconception of the proper length of time horizon. He argues that his three-year horizon concept makes sense on the basis of the correspondence between P_y as estimated by the three-year income correlation and the plotted slope. In other words, the estimation of P_y is a bootstrap operation (see Friedman p. 184). Friedman did get as far as noting that, "The size of the correlation between incomes in two successive years therefore provides some evidence on the importance of the permanent component in producing differences in measured income" (p. 184). But no exact statement of our equation (7) seems to appear anywhere in the *PIH* literature.

the structure of b_j's is the same for each individual, k the ratio of total national consumption to total national income in any year will be[9]

$$k = b_0 + b_1 + b_2 + \cdots + b_J \tag{8}$$

Because the r's are less than one, $E(\hat{b}_{0,c})$ must be less than k, the sum of the b's, which explains the cross-section slope (MPC) being less than the APC. This is the result that **F** and **M** worked so hard to derive in terms of "transitory" and "permanent" income.

II. *Quantitative Explorations of Cross-Sectional Budget Studies*

The formulation given above allows us to make some quantitative explorations which were not possible before. Almost all of **F**'s tests of the permanent-income hypothesis (PIH) were qualitative, being comparisons in which he examined the difference between groups. As an example; Blacks, who have a lower measured consumption ratio than Whites at the same measured income, also have a lower mean income, which accords with **F**'s hypothesis. But with the aid of the above analysis many of these qualitative tests can be examined quantitatively to see to what extent the permanent-income hypothesis (and a particular set of coefficients) explain the observed data. To be more specific, we may now test the extent to which the data are consistent with the proportionality hypothesis that MPC equals APC. To put it another way, we may reckon *how much* of the observed discrepancy between APC and MPC in cross-section (and time-series) studies may be explained by what J. Johnston (pp. 148–49) calls "errors in variables," or perhaps better, the omission of lagged variables from the analysis.

The meaning and importance of the proportionality hypothesis and its relationship to the rest of the **F** and **M** theorizing is complicated. Suffice it here to say that if proportionality is shown in budget cross-sections, it has some implications of its own (e.g., for long-run theorizing about secular stagnation, but not for short-run forecasting). It also would lend support to the most basic aspects of the **F** and **M** discussions, i.e., that people's consumption is affected by incomes in periods other than the current period. The support comes from the fact that if the influence of a tendency to *long-run* proportionality helps explain the data, it must show that income in other years does indeed affect short-run consumption. To turn it around, if income in other periods does not influence consumption in this period, a long-run analysis should be no closer to proportionality than the short-run analysis.

Before proceeding, we want to draw the reader's attention to the work of T. Mayer (1966), who recently carried out an ingenious related study.[10] He reasoned that for a large and varied sample of broad occupational groups, transitory income averages out to zero. Thus the mean measured income of a group can be taken to be its mean permanent income. If the "strict" proportionality hypothesis is correct, a plot of the mean measured-consumption, measured-income ratios for occupational groups with different mean incomes would pass through the origin. And if it does not go through the origin, the slope and height allows one to say *how much* of the difference between APC (for the society as a whole) and MPC is explained by the *statistical illusion* due to the permanent-income effect.

Mayer found that "... the results ... strongly reject the modified permanent income theory ... if one uses permanent income in place of measured income, only

[9] Or, $k = 1/L + b_0 + b_1 + b_2 + \ldots + b_J$ corresponding to (7a).

[10] See also Mayer's 1963 paper and Eisner.

about one-third of the difference between the average propensity and marginal propensity to consume disappears" (1966, pp. 1167 and 1170).

Now we turn to some examples of how the logic in Section I can be used to estimate the relationship between *APC* and *MPC*.

1. *The Slopes of Cross-Sectional Budget Studies*

The first example is that of determining how much of the difference between k and b_0, the average and short-run marginal propensities to consume, may be explained by the lagged effect of income (i.e., by the "permanent-income" effect), on the evidence of common budget studies. F estimated equation (3) from aggregate data on the assumption that the lag structure is exponential. F then chose those coefficients as the ones he thought appropriate for the effect of past (i.e., expected future) years' incomes. We shall work with the same coefficients without further ado for lack of any better set of weights to use.[11] This structure of weights has a central place in all the subsequent analysis of Section II. The reader should remember that the specific quantitative conclusions we reach in this section about the permanent-income and related hypotheses depend upon the weights we are using, which means that this sort of test is a test of *both* F's weights *and* the proportionality hypothesis together. Any failure of proportionality to appear may be caused by *either* lack of proportionality or wrong weights (or by the model in equation (6) being wrong, i.e., that factors other than past income affect current consumption through expected income, or that factors other than assets and expected income need to be in the consumption function, e.g., past consumption).

[11] The reader may be interested in the criticism of Friedman's weights by Holbrook.

If proportionality holds, then (considering equation (8)) the coefficients must sum to .88, Friedman's estimate of k (1957, p. 117). From Friedman's Table 15 (p. 147fn) the lag coefficients are roughly as follows (assuming a cut-off at seven years): $b_0=.31$, $b_1=.21$, $b_2=.14$, $b_3=.09$, $b_4=.06$, $b_5=.04$, $b_6=.03$.

In order to implement equation (7) we require the period-to-period correlation coefficients of income. One set is available from a study of Wisconsin taxpayers (as cited by Friedman 1957, p. 187): $r_{Y_TY_{T-1}}=.83$, $r_{Y_TY_{T-2}}=.78$, $r_{Y_TY_{T-3}}=.76$, $r_{Y_TY_{T-4}}=.71$, $r_{Y_TY_{T-5}}=.70$, $r_{Y_TY_{T-6}}=.69$. Now substituting directly into equation (7) we get

$$\begin{aligned} E(b_{0,c}) &\cong .31 + .83 \times .21 + .78 \times .14 \\ &\quad + .76 \times .09 + .71 \times .06 \\ &\quad + .70 \times .04 + .69 \times .03 \\ &\cong .75 \end{aligned}$$

This expected value of approximately .75 for $\hat{b}_{0,c}$ is the marginal propensity to consume *that we would expect to observe in a cross-section budget study if the permanent-income hypothesis is strictly correct* and if the b_j and r values used are correct.

We may now compare the .75 estimate to actually observed budget-study slopes. Empirical estimates of $\hat{b}_{0,c}$ range from .75 for the 1950 *U.S.* sample (Friedman 1957, p. 41)—which, taken together with the estimate of $E(\hat{b}_{0,c})$ above, would suggest that the permanent-income hypothesis explains all of the effect—downward to estimates that *average* perhaps .70. If we now take .70 as the working figure, our estimate of $E(\hat{b}_{0,c})$ as .75 explains almost three-quarters of the discrepancy between the cross-section *MPC* and the *APC*, if the latter is .88.

In any case, this approach suggests that the permanent-income hypothesis explains considerably more than the one-

third of the effect as estimated by Mayer.[12] On the other hand, to the extent that variations in income from year to year are perceived as one-time changes, e.g., windfalls, one would expect the increment to have relatively *little* effect on current consumption, by **M**'s reasoning.

2. *An Occupational-Stratification Test*

Friedman suggested the following test:

> Suppose a regression were computed for a broad group of consumer units, say a sample of all units in the United States, and the corresponding elasticity estimated. Suppose this broad group were broken down into subgroups, say by the communities in which they reside, and separate regressions computed for each community. An appropriately weighted average of the corresponding elasticities should then be smaller than the elasticity for the group as a whole, and smaller by an amount calculable from the income data for the separate communities. The classification by communities eliminates one source of variability in permanent components, and so should reduce the variance of permanent components and hence the elasticity. This process can be continued. For each community the groups can be classified by occupation; within occupation by education; within education by age and family size, and so on. At each stage in the hierarchy the average of the elasticities of the different groups should be lower than the elasticity for the broader group of which they are part. As the groups are more and more rigorously defined, the elasticity should approach zero. [1957, p. 216]

Houthakker performed such a test and found that "... the number of unfavorable cases (those in which the elasticity for the sub-group is larger than for the group as a whole) exceeds that of the favorable cases ... the combined impression is very damaging to the permanent-income hypothesis" (p. 403). Houthakker's analysis contained an error which, upon correction (see Eisner), resulted in a reversal of these findings, toward a "modified" permanent income theory. The (correct) logic of Friedman's suggested test is easily demonstrated with algebra on the least squares regression framework.

Viewed through our equation (7), yet another interpretation is available. Writing

$$E(\hat{b}_{0.c};\ \text{steelworkers}) \cong b_0 + b_1 r_{Y_T Y_{T-1}},\ \text{steelworkers} + b_2 r_{Y_T Y_{T-2}},\ \text{steelworkers} \cdots \tag{9}$$

and

$$E(\hat{b}_{0.c};\ \text{entire population}) \cong b_0 + b_1 r_{Y_T Y_{T-1}},\ \text{entire population} + b_2 r_{Y_T Y_{T-2}},\ \text{entire population} \cdots \tag{10}$$

we see that Houthakker's "favorable case" argument implies that the correlation coefficients of current income with its lags are lower for an occupational group, e.g., steelworkers, than for the population as a whole. If so, $E(\hat{b}_{0.c};$ steelworkers) would be lower than $E(\hat{b}_{0.c};$ population). But this is not true by definition, nor is it always true in fact. Consider the following estimates of $r_{Y_T Y_{T-1}}$ gleaned from Friedman's collection:[13] population as a whole, estimated from Wisconsin taxpayers, $r_{Y_T Y_{T-1}} = .83$; professions, $r_{Y_T Y_{T-1}} = .85$; clerical and sales, $r_{Y_T Y_{T-1}} = .88$; farm operators, $r_{Y_T Y_{T-1}} = .68$. Apparently there are some subgroups for which one would expect, *ceteris paribus*, a higher $\hat{b}_{0.c}$ than for the population as a whole, and hence Houthakker's test does not prove anything as it stands.[14] Eisner's reanalysis of

[12] Estimates using shorter-horizon weights such as those of Houthakker-Taylor (1966) or Liviatan would give results more in Mayer's direction.

[13] From Tables 19 (p. 180) and 20 (p. 195).

[14] Houthakker's findings may also have been affected by the fact that his observations were for income groups, sometimes as few as five groups, rather than for continuously-graded incomes (p. 402). Such grouping tends

similar data, tending to reverse Houthakker's findings, is based on much the same logic we have used in the comparison of equations (9) and (10).

Now we may consider the extent to which different *MPC*'s for different groups may be explained by the permanent-income hypothesis.[15] The two groups for which we have most of the necessary data are farmers and urban dwellers (see Table 1). If we apply Friedman's estimated lags just as before we get

$$E(\hat{b}_{0,c};\text{ farmers})$$
$$\cong .31 + .68 \times .21 + .63 \times .14 + .61 \times .09 + .56 \times .06 + .55 \times .04 + .54 \times .03$$
$$\cong .67$$

and

$$E(\hat{b}_{0,c};\text{ urban families})$$
$$\cong .31 + .83 \times .31 + .80 \times .14 + .76 \times .09 + .74 \times .06 + .72 \times .04 + .71 \times .03$$
$$\cong .76$$

So $E(\hat{b}_{0,c};\text{urban families}) - E(\hat{b}_{0,c};\text{farmers}) \cong .76 - .67 = .09$, whereas the observed $(\hat{b}_{0,c};\text{ urban families} - \hat{b}_{0,c};\text{ farmers}) = .79 - .57 = .22$. In this test the permanent-income effect explains a little less than half the observed difference.

Several comments are in order about this test. First, on the positive side, the fact that the computations use many "guesses" for the required correlation coefficients does not seem to be too important. As the reader may easily verify, the comparison is quite robust to different

TABLE 1—LAG COEFFICIENTS OF CORRELATION IN INCOME: FARM AND URBAN FAMILIES*

	Farm Families	*Urban Families*
Marginal propensity to consume, 1941	.57[a]	.79[a]
$r_{Y_TY_{T-1}}$	.68[b]	.83[c]
$r_{Y_TY_{T-2}}$	[.63]	[.80]
$r_{Y_TY_{T-3}}$	[.61]	[.76]
$r_{Y_TY_{T-4}}$	[.56]	.74[d]
$r_{Y_TY_{T-5}}$	[.55]	[.72]
$r_{Y_TY_{T-6}}$	[.54]	[.71]

* The data in brackets are *guesses* by the authors. The only basis for the guesses was the distribution of Wisconsin taxpayer serial correlations mentioned earlier. See the text for some reassurance about the results of these guesses.

[a] Friedman (1957), p. 41.
[b] Friedman (1957), p. 195 for 1947–48.
[c] Friedman (1957), p. 187 for 1947–48.
[d] Friedman (1957), p. 189.

"guestimates." There are, however, valid objections to the comparison, as our referee has pointed out. One objection is that different groups may be expected to have different lag weights, and Friedman's aggregate lag weights cannot be taken to be good proxies for either the lag structure of farmers or urban families. A second objection is that the consumption behavior of farmers may be fundamentally different from the behavior of the rest of the population. We have decided to include this test, nevertheless, because Friedman considered the comparison worth making in 1957.

3. *Income and Consumption Data for Two Years Pooled*

Lawrence R. Klein found for *U.S.* data that "the savings-income ratio of twenty-four months was less sensitive to income classifications than that of annual data" (p. 221), which is in the direction predicted by Friedman's hypothesis. That is, Klein found that the savings-income ratio was less strongly correlated with the level of 24 months' income than with data based on a 12-months period.[16] Now let us see

to reduce errors in variables, which is exactly the nature of the permanent-income hypothesis logic (see Johnston, p. 148–49). His results may also be affected by different lag structures, i.e., more or less "foresight" in different groups. (See Mayer (1966), pp. 1173–74, and our later discussion.)

[15] This quantity, Friedman says, is ". . . calcuable from the income data for the separate communities" (extracted from quote given above). But as Houthakker noted, Friedman "does not say how to calculate it." (p. 992).

[16] This finding was reversed for British data (Klein and Liviatan, p. 156).

what equation (7) suggests about $\hat{b}_{0,c}$ if the period of analysis is two years instead of one. Writing (3) for C_{T+1} and adding to C_T gives the structure

$$
\begin{aligned}
(11)\quad & C_{i,T} + C_{i,T+1} \\
& \quad = b_0(Y_{i,T+1} + Y_{i,T}) \\
& \qquad + b_1(Y_{i,T} + Y_{i,T-1}) \\
& \qquad + b_2(Y_{i,T-1} + Y_{i,T-2}) + \cdots
\end{aligned}
$$

Assuming that we run $(C_{i,T}+C_{i,T+1})$ on $(Y_{i,T}+Y_{i,T+1})$, the bias form is

$$
\begin{aligned}
(12)\quad & E(\hat{b}_{0,c};\ \text{two-year periods}) \\
& \cong b_0 + b_1 r_{T[T+(T+1)]Y[T+(T-1)]} \\
& \qquad + b_2 r_{Y[T+(T-1)]Y[(T-1)+(T-2)]} + \cdots
\end{aligned}
$$

Now supposing that $\sigma^2_Y = \sigma^2_{Y_{T-1}}$, etc., and cov (Y_T, Y_{T-1}) = cov (Y_{T-1}, Y_{T-2}), etc., the required correlation coefficients in (12) may be calculated in terms of their available constituent parts, as

$$
\begin{aligned}
(13)\quad & r_{Y[T+(T+1)]Y[(T-J)+(T-J-1)]} \\
& \qquad = \frac{r_{Y_T Y_{T-J}} + 2r_{Y_T Y(T-J-1)} + r_{Y_T Y(T-J-2)}}{2(1+r_{Y_T Y_{T-1}})}
\end{aligned}
$$

Carrying out the computations suggested by (11) and (12) with the data at hand gives an expected value of .76 for $\hat{b}_{0,c}$ from two-year data under the permant-income hypothesis, as compared to the .75 expected value calculated previously for annual information. This jibes with Klein's findings, though his article does not give a two year *MPC* for quantitative comparison, and he gives the 12 month regression for only one of the two years. (The other year's results could easily be such as to reverse Klein's finding.) As one would expect, further aggregation of time periods will theoretically produce even higher expected *MPC*'s to a limit of *APC*.

III. *Quantitative Explorations of Time-Series Consumption Functions*

"I have not been able to find any way to predict the quantitative relation between the two [time-series and budget studies]," *F* wrote (1957, p. 135). He also said he was unable to predict the quantitative differences between time-series covering periods of different lengths, the supposed reason being that ". . . a steady secular trend is not the only factor producing differences in permanent income . . ." (p. 127). But we are now in a position to do both these tasks. We have seen that the method above predicts the outcome that may be expected from a cross-section sample of a population with given parameters. The same method allows us to state what outcome may be expected of a conventional time-series analysis of the consumption function that omits lagged variables, by substituting directly into equation (7). The estimates of the correlation coefficients are now the various *serial correlations*. The serial correlations $\hat{r}_{Y_T Y_{T-1}}$ and $\hat{r}_{Y_T Y_{T-2}}$ may be high for very long time-series, e.g., .97, .91, .85 and .79 for the first five lags for per capita incomes for the period 1897 to 1949, using standard *U.S.* historical data. On the other hand, these serial correlations tend to be much lower for short time-series during periods of cyclical variation, e.g., .78, .40, .04 and −.36 for the successively more distant lags for the period 1923 to 1940. So this immediately explains the finding that the estimated *MPC* is higher from long time-series regressions than short ones. This also enables us to estimate how much of the discrepancy between k and $\hat{b}_{0,t}$, the time-series estimate of b_0, is explained by the omission of lagged variables, in exactly the same manner as was done for cross-sections above.

Assuming that the structure (7) holds for time-series aggregate (or average) variables, let us calculate the expected *MPC* for the two examples, 1923–1940 and 1897–1949:

1923–1940

$$E(\hat{b}_{0,t}) \cong .31 + .78 \times .21 + .40 \times .14 + .04 \times .09 - .36 \times .06 = .51$$

1897–1949

$$E(\hat{b}_{0,t}) \cong .31 + .97 \times .21 + .91 \times .14 + .85 \times .09 + .79 \times .06 + .73 \times .04 + .67 \times .03 = .80$$

Now we may compare the above to empirical MPC estimates for these periods using deflated per capita series. For 1897–1949: .70 (Goldsmith, durables); .74 (Goldsmith, nondurables); .91 (Goldsmith, excluding the war years). For 1923–1940: .36 (Goldsmith, nondurables); .79 (Ferber, durables). The variability in the observed estimates is sufficiently great that it prevents our asking what proportion of the difference between APC and MPC is accounted for by omitted lagged variables (i.e., the permanent-income effect). Perhaps the best we can say is that the sets of observed MPC's *bracket* the expected value for both 1897–1949 and for 1923–1940. Besides suggesting that even a "strict" proportionality hypothesis is not thereby disproven, this suggests that F's structure of coefficients might be sound.

The difference in results between the cross-sectional budget studies and the time-series regressions requires explanation. The best explanation that we can offer is that (a) there is *some* relative-income effect, as Duesenberry suggested; and/or (b) there is a secular shift upwards in APC, partly as a function of the introduction of new goods. This explanation is not very parsimonious, but we believe that it is the most plausible reconciliation of the (adjusted) results from the two observational techniques.

IV. *The Lag Coefficients Implied by the Proportionality Hypothesis and the Available Data*

Until now we have asked to what extent the variation in income from year to year, together with a given set of lag coefficients, explains the discrepancy between the observed MPC in cross-section budget studies and the APC. Now let us take another tack and ask *how great the lags would have to be* to explain the discrepancy between MPC and APC, given the observed year-to-year variation in income (and the basic model in equation (6)).

We may do this by setting $E(\hat{b}_{0,c})$ in equation (7) equal to the observed MPC in a given cross-section or time-series study, inserting estimates of $r_{Y_T Y_{T+j}}$, and solving for the b_j lag weights, assuming a geometric structure, say, as before. Let us first do this for some examples in the range of observed cross-section MPC's, starting with .75 and working downward: we begin with equation (7), putting $E(\hat{b}_{0,c}) = .75$ and utilizing the lag correlations of Section II.1. This gives $.75 = b_0 + .83b_1 + .78b_2 + .76b_3 + .71b_4 + .70b_5 + .69b_6$, where $b_1 = b_0\beta$, $b_2 = b_0\beta^2$, $b_3 = b_0\beta^3$, etc., $b_0 \cong .88(1-\beta)$, and $\sum_{j=0}^{\infty} b_j = .88$ (which is k as estimated by Friedman). We then wish to solve for the structure of the b's, or to put it another way, to solve for β.[17] For $E(\hat{b}_{0,c}) = .75$, $\beta = .67$ and the b's are $b_0 = .31$, $b_1 = .21$, $b_2 = .14$, $b_3 = .09$, $b_5 = .04$ and $b_6 = .03$, as was seen earlier. For $E(\hat{b}_{0,c}) = .70$, $\beta = .65$, and the b's are $b_0 = .31$, $b_1 = .20$, $b_2 = .13$, $b_3 = .08$, $b_4 = .05$, $b_5 = .03$ and $b_6 = .02$. For $E(\hat{b}_{0,c}) = .65$, $\beta = .71$, and the b's are $b_0 = .26$,

[17] Which amounts to solving a J-order polynomial equation in β. Considering the reverse (bias) problem, with a geometric series of weights the bias in $\hat{b}_0$ depends on the correlation structure and β only. A maximum bias can be calculated by setting all correlations equal to one, which results in the simple formula: max $[E(\hat{b}_0) - b_0] = b_0\beta/(1-\beta)$. (See also Grunfeld p. 400.)

$b_1=.18$, $b_2=.13$, $b_3=.09$, $b_4=.06$, $b_5=.05$, and $b_6=.03$.[18] Whether these structures of weights are reasonable is up to the reader to judge on the basis of his general knowledge of consumption behavior; there is apparently no other test. What we can say, since these weight structures differ from Friedman's, is that if we observe cross-section slopes that center around .65, for instance, he is either wrong on proportionality or on his weights or both.

Finally we may find lag structures that would be consistent with observed *MPC*'s for cross-sections of farm and urban families (see Section II.2). For farm families we have $.57=b_0+.68b_1+.63b_2+61b_3+.56b_4+.55b_5+.54b_6$, which gives $\beta=.81$, $b_0=.17$, $b_1=.14$, $b_2=.11$, $b_3=.09$, $b_4=.07$, $b_5=.06$, and $b_6=.05$. For urban families, $.79=b_0+.83b_1+.80b_2+.76b_3+.74b_4+.72b_5+71b_6$, with $\beta=.49$, $b_0=.45$, $b_1=.22$, $b_2=.11$, $b_3=.05$, $b_4=.03$, $b_5=.01$, and $b_6=.01$.[19]

For time-series it is not really necessary to perform the same exercise. The sets of actual estimates for the periods 1897–1949 and 1923–40 bracket the expected *MPC*'s, and it seems not unreasonable that the variation might be due to data and other research-procedural matters. So a lag structure like the one estimated by F may well be sufficient to bring the observed *MPC* in non-lagged regressions into line with *APC*, and hence lends support to that structure of lags for the time-series analysis.

What conclusions can be drawn from lag structures which are consistent with the idea that *MPC* is equal to *APC* in cross-sectional budget studies? Our judgment is that even if aggregate *MPC* equals aggregate *APC* over time as income grows, the true family *MPC* is not equal to *APC*, i.e., the proportion of income saved is larger for families with higher (long-run) incomes. The discrepancy between *MPC* and *APC* is not as great as a single one-period cross-section suggests, however, the errors-in-variables phenomenon being responsible for a considerable part of the discrepancy. But whether the discrepancy is larger or smaller, it need not have any aggregate implications, as noted earlier.

V. *Summary*

A simple device based on the notion of specification bias due to omitted lagged variables was used to show why observed propensities to consume in one-period cross-section studies and in simple time-series regressions may differ from the "true" propensity to consume. Then the device was used to crudely estimate how much of the discrepancy between observed *MPC*'s and *APC*'s may be accounted for by this specification bias. This exercise requires the assumption of some set of lag coefficients, however. Hence the exercise jointly tests the proportionality hypothesis and the set of coefficients used. Last, we discussed the set of lag coefficients that would be required to make the observed *MPC*'s jibe with the observed *APC*.

REFERENCES

A. Ando and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55-84.

R. Eisner, "The Permanent Income Hypothesis," *Amer. Econ. Rev.*, Dec. 1958, *48*, 972-90.

M. Friedman, *A Theory of the Consumption Function*. Princeton 1957.

———, "Windfalls, the 'Horizon', and Related Concepts in the Permanent-Income Hypothesis," in *Measurement in Econom-*

[18] These two cases each produced an additional real root which was negative. The sums of coefficients are .82 and .80, respectively, the discrepancy from .88 due to two-digit rounding and the approximation $b_0 \cong .88(1-\beta)$.

[19] In these two cases the coefficient sums are .69 and .88. The approximation $b_0 \cong .88(1-\beta)$ which is exact for an infinite geometric series can be expected to be poor for β close to one at a cut-off of six terms.

ics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld, Stanford 1963, pp. 3-28.

Y. Grunfeld, "The Interpretation of Cross-Section Estimates in a Dynamic Model," *Econometrica,* July 1961, *29,* 397-404.

R. S. Holbrook, "The Three-Year Horizon: An Analysis of the Evidence," *J. Polit. Econ.,* Oct. 1967, *75,* 750-54.

H. S. Houthakker, "The Permanent Income Hypothesis," *Amer. Econ. Rev.,* June 1958, *48,* 396-404.

——— and L. D. Taylor, *Consumer Demand in the United States, 1929-1960.* Harvard 1966.

J. Johnston, *Econometric Methods.* New York 1963.

L. R. Klein, "Statistical Estimation of Economic Relations from Survey Data," in L. Klein, ed., *Contributions of Survey Methods in Economics,* New York 1954, Ch. 5.

——— and N. Liviatan, "The Significance of Income Variability on Savings Behavior," *Bull. Oxford Univ. Inst. Econ. Statist.,* May 1957, *19,* 151–60.

N. Liviatan, "Estimates of Distributed Lag Consumption Functions from Cross Section Data," *Rev. Econ. Statist.,* Feb. 1965, *47,* 44–53.

E. Malinvaud, *Statistical Methods of Econometrics,* Amsterdam 1966.

T. Mayer, "The Permanent Income Theory and Occupational Groups," *Rev. Econ. Statist.,* Feb. 1963, *45,* 16–22.

———, "The Propensity to Consume Permanent Income," *Amer. Econ. Rev.,* Dec. 1966, *56,* 1158–77.

F. Modigliani, "The Life Cycle Hypothesis of Saving, The Demand for Wealth, and the Supply of Capital," *Soc. Res.,* summer 1966, *33,* 160–217.

——— and R. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. Kurihara, ed., *Post-Keynesian Economics,* Rutgers 1954, pp. 383–436.

——— and A. Ando, "The 'Permanent Income' and the 'Life Cycle' Hypothesis of Saving Behavior: Comparisons and Tests," in *Proceedings of the Conference on Consumption and Saving,* Vol. 2. Philadelphia 1960.

——— and ———, "Test of the Life Cycle Hypothesis of Saving," *Bull. Oxford Univ. Inst. Econ. Statist.,* May 1957, *19,* 99–124.

J. L. Simon, "A Reconciliation of the Life-Cycle and Permanent Income Theories," Univ. Illinois, mimeo 1969.

Absenteeism and the Overtime Decision

By RONALD G. EHRENBERG*

Upon reading the congressional hearing on the *Overtime Pay Penalty Act of 1964*, one cannot fail to be impressed by the emphasis that management places on absenteeism as a primary cause of overtime. The argument given is basically quite simple: Large firms, it is claimed, attempt to account for absenteeism by hiring standby workers; however because of the stochastic nature of the absentee rate, it is impossible for them to have replacements always available. Hence overtime must be worked by existing employees in order to meet production schedules. One concludes from this argument that the randomness of absenteeism is the cause of overtime. If the absentee rate were known with certainty, then management could take account of it without recourse to additional overtime.[1]

In this note we challenge this conclusion and argue that a rational economic response to a certainty absentee rate involves *increasing* the amount of overtime worked per man, while the effect on the level of employment is ambiguous. Furthermore we claim that a stochastic absentee rate leads to a larger optimal employment stock and, in at least one special case, on average, to a *smaller* amount of overtime worked per man than in the certainty absentee rate case. Crucial to our argument is the observation that many of the labor costs which we classify as "fixed" must be paid by the employer even when an employee is absent, but overtime hourly wage payments need not be made to absentees.

Section I presents the structure and assumptions of a simple static model, similar to that found in Sherwin Rosen's paper, when absenteeism is zero. The next section generalizes this model to include a certainty absentee rate. In Section III, we consider the case of a stochastic absentee rate and compare the optimal solution with the results of the previous section. The final section briefly summarizes the results.

I

Given the level of output to be provided, the level of technology, and the flow of capital services, a neoclassical production function can be inverted to determine a unique required flow of labor services. The firm decision problem is then to choose that combination of men and hours per man which will produce that flow, and which will minimize its labor costs. Symbolically, the problem is to:

$$\text{(1)} \quad \text{minimize } w_1M + (r+q)TM + w_2M\bar{H} + w_3M(H-\bar{H})$$

* Department of economics, Northwestern University. Without implicating them for what remains, I am grateful to Dale Mortensen and George Delehanty for comments on an earlier draft. Finis Welch suggested the approach taken in Figure 1, that both generalized the results and simplified the presentation of the original note, which was based solely on the Cobb-Douglas function. Research was supported by a grant from the Manpower Administration of the U.S. Department of Labor under the Manpower Development and Training Act of 1962, as amended.

[1] This conclusion and the argument that follows neglect the indivisibilities inherent in small firms, union rules concerning the existence of stand-by workers, and the heterogeneity and scarcity of skilled labor. While all these factors tend to limit the employment of standby workers they lead us to expect that overtime is positively related to the absentee rate.

(2) subject to $L = F(M, H)$

where (a) $F_1, F_2 > 0$

(b) $(2F_1F_2F_{12} - F_1^2F_{22} - F_2^2F_{11}) - (2F_1F_2^2/M) > 0$

Here w_1 represents those employment costs per man which are fixed in the sense of being independent of the exact number of hours that each employee works. These include the costs for such items as paid vacations, paid holidays, private welfare and insurance plans, and many legally required insurance payments.[2] Some of these costs are annual, others monthly, still others weekly; the assumption here simply being that the employer imputes them to himself on a weekly basis. The next term represents what Rosen and M. Ishaq Nadiri have called the "user cost of labor." T represents the once-over turnover and investment cost per man of hiring and training workers. If these costs are financed by borrowing, they must be discounted by the interest rate, r, and also adjusted for expected replacement costs by the quit rate, q. Assuming that equilibrium occurs in the overtime region, the wage costs per man are the wage rate, w_2, times the maximum number of hours per man payable at straight-time wages, $\bar{H}$, plus the overtime wage rate, w_3, times the number of overtime hours per man, $(H-\bar{H})$.

The constraint (2) asserts that the flow of labor services, L, is a function of the number of men employed, (M), and the number of hours per man, (H). For a number of reasons discussed by Martin Feldstein, it is inappropriate to specify the labor input as being equal to the number of man-hours worked, (MH). Here we assume only that the marginal contribution to labor services of each input is positive over the relevant region, and that the necessary condition (2b) for the optimizing problem to have a solution is met. This condition requires that the marginal rate of substitution of men for hours be a decreasing function of hours, the usual convex isoquant assumption.[3]

Upon minimizing (1) subject to (2), these assumptions lead to an equilibrium combination of men and hours (M^*, H^*) which is a function of all of the parameters in the model. In particular, it is important for later use to note that an increase in the component of cost independent of hours $(w_0 = w_1 + (r+q)T)$ increases the marginal cost of labor through additional employment relative to the marginal cost of labor through added hours per worker. Consequently a substitution of overtime hours for employment would occur.

(3) $\partial H^*/\partial w_0 > 0, \quad \partial M^*/\partial w_0 < 0$

II

Suppose we now introduce a certainty absentee rate into the model. That is, the firm knows that at any given day only the fraction "a" of its employees will be in attendance. Then its appropriate labor input function becomes

(2′) $$L = F(aM, H).$$

The inclusion of an absentee rate also modifies the cost function, but in a nonsymmetric way. In particular, the capitalized turnover costs must be paid regardless of whether an employee works on any given day (or week). Similarly many of the fixed employment costs such as health insurance, pension coverage, vacation pay, and unemployment compensation insurance are independent of the

[2] That the magnitude of these costs is not small, can be seen by consulting *B.L.S.* publications such as Bulletin 1428.

[3] Actually condition (2b) is stronger than the requirement of diminishing marginal rate of substitution between factors because a constant budget cost curve is not linear when viewed in (M, H) space. See Rosen (p. 515) for a graphical illustration of this point.

employee's attendance on any particular day.[4] For simplicity we initially assume that they are all independent of attendance. Finally wage costs, for the most part, are paid only to workers actually working, hence the appropriate cost function becomes

$$(1') \quad [w_1 + (r + q)T]M + w_2 aM\overline{H} + w_3 aM(H - \overline{H}).$$

Through a simple transformation of variables, it is easy to see directly what the effect of the certainty absentee rate is on the equilibrium values of men and hours. Let $A = aM$, the number of employees actually working on a given day. Then rewriting (1′) and (2′) the firm seeks to

$$(1'') \quad \text{minimize } [(w_1 + (r + q)T)/a]A + w_2 A\overline{H} + w_3 A(H - \overline{H})$$

$$(2'') \quad \text{subject to } L = F(A, H)$$

Obviously, in terms of the optimal A, H combination, a decrease in a (an increase in the absentee rate $1-a$) has the same effect as an increase in any of the other components of costs that are independent of hours, w_0. That is, an increase in absenteeism increases the marginal cost of labor through additional workers in attendance, relative to the marginal cost of labor through additional overtime. Consequently

$$(3') \quad \frac{\partial H^*}{\partial(1-a)} > 0, \qquad \frac{\partial A^*}{\partial(1-a)} < 0$$

An increase in absenteeism causes a substitution of additional overtime per man for workers in attendance. In general, however, we cannot predict the effect on the equilibrium employment stock. Since $A = aM$ we know that

$$(4) \quad \frac{\partial M^*}{\partial(1-a)} = \frac{1}{a}\frac{\partial A}{\partial(1-a)} + \frac{M}{a}$$

While an increase in absenteeism (a decrease in the attendance rate) causes a substitution effect which tends directly to decrease employment (the first term in (4)), it also causes a scale effect since more employees are now required to attain a given level of workers in attendance on any day. Because the substitution and scale effects work in opposite directions, it is impossible to predict what the net effect of a certainty absentee rate is on the number of employees.

Note that the above results will continue to hold even if some of the fixed employment costs, such as daily travel expenses, need not be paid to absentees. Similarly they will hold if some (or all) absentees receive sick leave payments. All that we require is that absentees do *not* receive any pay for overtime hours that they may have been scheduled to work and that some of the employment costs which are independent of hours worked are also independent of attendance.[5]

III

Instead of being known with certainty, we now assume that the attendance rate is a random variable with a probability density $p(a)$ that is symmetric[6] around the mean value, $E(a)$. As in the recent works of Kenneth Smith and Michael Rothschild, we assume that the firm faces a two-stage decision process. The employer

[4] This last statement should be qualified. Often no holiday pay is received unless the employee works the days directly before and after a paid holiday. Similarly unless a minimum number of days are worked, the employee is ineligible for vacation pay and pension credit. Also unemployment compensation insurance costs are man-hour (not man) related unless the employee's annual income is above a certain level. Finally, if the employee is "fired for cause," all of these obligations cease.

[5] This proposition which can be rigorously proved has been omitted for brevity. For notational simplicity we ignore these generalizations in what follows since they in no way effect the results of the next section.

[6] Obviously we also require $P_{(a)} \geq 0$,

$$\int_0^1 P(a) = 1, \qquad \int_0^1 aP(a) = E(a).$$

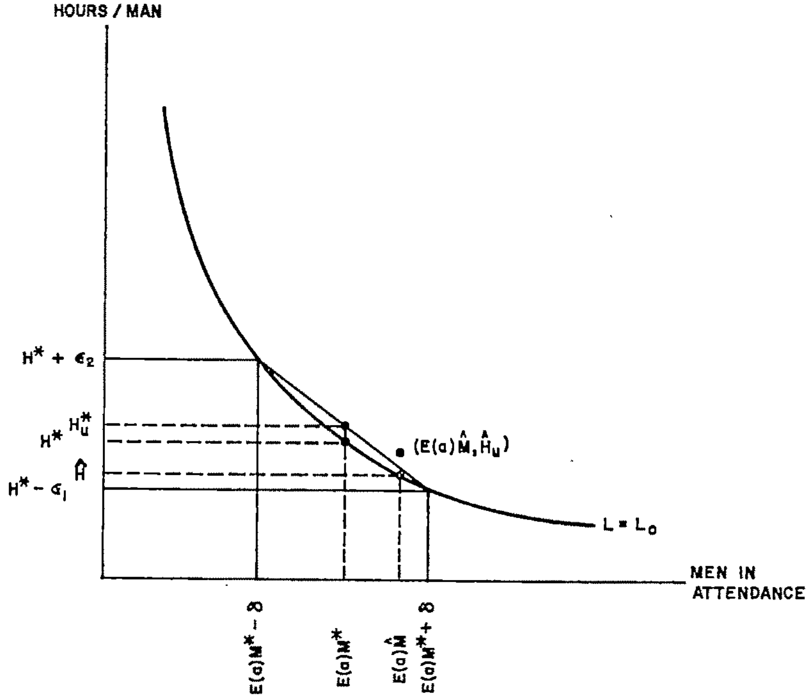


FIGURE 1

must choose the employment stock, M, each period before the actual realized value of the attendance rate is known. Once the attendance rate is observed, then hours per man, H, is determined from the labor services requirement (2′) in order to meet production schedules. The two-stage process seems particularly relevant in this context, in light of our discussion in the opening paragraph.

It is easy to illustrate graphically, that due to the assumption of diminishing marginal rate of substitution between workers in attendance and hours per worker, that random fluctuations of absenteeism about a given mean level serves to *increase* the optimal stock of employees.

Referring to Figure 1, the contour $L = L_0$ indicates the various combinations of workers in attendance and hours per worker which yield the required level of labor services, L_0. Suppose that when absenteeism was nonstochastic that the equilibrium was given by the combination $(E(a)M^*, H^*)$. Now if the attendance rate is stochastic, symmetric fluctuations in worker attendance about $E(a)M^*$ (of size δ for example) will require compensating fluctuations in hours per worker. However, because of the assumed diminishing marginal rate of substitution, negative deviations in attendance require more than proportional increases in hours, while positive deviations require less than proportional increases (in Figure 1, $\epsilon_2 > \epsilon_1$). As a result, if the attendance rate is distributed symmetric about its mean, average hours associated with M^* in the stochastic case (H_u^*) will exceed the level of the certainty case (H^*). If we assume that the objective of the firm is to minimize the expected value of labor costs, then since $(E(a)M^*, H^*)$ represented the point at which the extensive (employee) marginal labor cost was equated to the intensive (hours/worker) marginal labor cost, we know that $(E(a)M^*, H^*)$ cannot be an equilibrium point. Effectively the in-

tensive marginal cost has increased relative to the extensive marginal cost; for a given number of men, a larger number of hours per man are required on average. Consequently, it is optimal to *increase* the number of employees to some level $\hat{M} > M^*$.

Associated with this employment level ($\hat{M}$), there is an optimal level of the expected value of hours per man ($\hat{H}_u$). Due to the stochastic absentee rate and assumption of diminishing marginal rate of substitution between factors, $\hat{H}_u$ is greater than the level of hours per man that would be associated with $\hat{M}$ in the nonstochastic case ($\hat{H}$). While it is certainly true that $\hat{H} < H^*$ and $\hat{H}_u < H_u^*$, the relevant comparison is between $\hat{H}_u$ and H^*. That is, would we observe on average, a greater or smaller level of overtime per man in the stochastic or nonstochastic case? Figure 1 does not give us sufficient information to answer this question and indeed we have been unable to compare these terms for the general class of labor input functions in (2). It appears that the comparison will depend upon properties of the labor input function, such as the elasticity of substitution between the factors, which determine the shape of the isoquant $L = L_0$.

If we consider the special case of the Cobb-Douglas labor input function, we can however uniquely determine the relationship between equilibrium overtime hours per man in the certainty and stochastic cases. That is, we assume

$$(2''') \qquad L = (aM)^{\alpha} H^{\beta}, \quad \alpha > \beta^{7}$$

In addition to its analytic convenience, we may also justify this function's use because it has been employed with success in recent empirical work by Martin Feldstein and Nadiri and Rosen. For this particular function, the solution for equilibrium hours in the case of a certainty absentee rate of $E(a)$ becomes

$$(5) \qquad H_c^* = \left[\frac{w_0 + (w_2 - w_3)\overline{H}E(a)}{w_3 E(a)}\right] \cdot \left[\frac{\beta}{\alpha - \beta}\right]$$

In the stochastic case, the firm knows that once it chooses an employment level M, hours per man will be uniquely determined by the value of the absentee rate that actual obtains in the period, i.e.

$$(6) \qquad H = L^{1/\beta}(aM)^{-\alpha/\beta}$$

We assume that the firm will attempt to choose M, conditional on the value of H in (6), to minimize the expected value of its labor costs. Symbolically, substituting (6) into (1′), the firm seeks to

$$(7) \qquad \underset{M}{\text{minimize}}\ E[w_0 M + (w_2 - w_3)\overline{H}aM + w_3 L^{1/\beta}(aM)^{(\beta-\alpha)/\beta}]$$

The necessary condition for this unconstrained minimization problem is that

$$(8) \qquad E\left[w_0 + (w_2 - w_3)\overline{H}a + \left(\frac{\beta - \alpha}{\beta}\right) w_3 L^{1/\beta} a^{(\beta-\alpha)/\beta} M^{-\alpha/\beta}\right] = 0$$

or that

$$(9) \qquad M^{-\alpha/\beta} = \frac{[w_0 + (w_2 - w_3)E(a)\overline{H}]}{w_3 L^{1/\beta} E[a^{(\beta-\alpha)/\beta}]}\left(\frac{\beta}{\alpha - \beta}\right)$$

Substituting (9) into (6) when the attendance rate takes its mean value $E(a)$ determines the optimal expected level of hours per man in the stochastic case

$$(5') \qquad H_u^* = \left[\frac{w_0 + (w_2 - w_3)E(a)\overline{H}}{w_3 E[a^{(\beta-\alpha)/\beta}]}\right][E(a)]^{-\alpha/\beta} \cdot \left(\frac{\beta}{\alpha - \beta}\right)$$

The optimal expected level of hours per man in the certainty (H_c^*)and stochastic (H_u^*) cases may now be directly compared. Dividing (5) by (5′) yields that

[7] The requirement $\alpha > \beta$ is the second order necessary condition corresponding to (2b) in the general case. This condition will be crucial to what follows.

$$\frac{H_o^*}{H_u^*} = \frac{E[a^{(\beta-\alpha)/\beta}]}{[E(a)]^{(\beta-\alpha)/\beta}} \tag{10}$$

Since $\alpha > \beta$ is a necessary condition for our solution in (5) to be a relative minimum, we can let $\alpha = K\beta$, where $K > 1$, and obtain

$$\frac{H_o^*}{H_u^*} = \frac{E(a^{1-K})}{[E(a)]^{1-K}}, \quad \text{or rewriting}$$

$$\left(\frac{H_u^*}{H_o^*}\right)^{1/(K-1)} = \frac{[E(a^{1-K})]^{1/(1-K)}}{E(a)} \tag{11}$$

Using Holder's inequality, it can be shown that the right-hand side of (11′) is always less than unity and hence[8]

$$H_u^* < H_o^*. \tag{12}$$

In this case the expected level of overtime per man is *less* than when the absence rate is known with certainty.

IV

Summarizing our results briefly, contrary to popular belief, it is not always the stochastic nature of absenteeism which is responsible for increased overtime hours per man above the zero absentee level. This is due to the fact that a certainty absentee rate modifies the labor cost function in a nonsymmetric way so as to increase the marginal cost of labor purchased through additional workers relative to the marginal cost of labor purchased through increased hours per man. Although a certainty absentee rate causes this substitution of hours per worker for workers, the net effect on the employment stock is ambiguous since the scale effect of increased absenteeism tends to increase employment.

A stochastic absentee rate tends to increase the optimal employment stock above the certainty absentee rate level. While the effect on the expected level of hours per man has not been determined in general, for the special case of the Cobb-Douglas labor input function it is shown that the optimal level decreases. That is, observed overtime hours per man would be lower in the stochastic than nonstochastic case.

[8] See G. H. Hardy, et al., (pp. 134–45) for a statement and proof of Holder's inequality and relevant corollaries. In particular it is shown that for t any real numbers, $[E(x^t)]^{1/t}$ is a monotonically increasing function of t. Note that if we define variability more generally in terms of a "mean preserving spread," then M. Rothschild and J. Stiglitz have shown that for any convex function $f(a)$, the expected value $E(f(a))$ rises when the variability increases. Since a^{1-K} is convex for $K > 1$, it immediately follows from (11′) that an increase in the variability of absenteeism decreases the optimal level of overtime per man.

REFERENCES

M. S. Feldstein, "Specification of the Labor Input in the Aggregate Production Function," *Rev. Econ. Stud.*, Oct. 1967, *34*, 375–86.

G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge 1934, pp. 134–45.

M. I. Nadiri and S. Rosen, "Interrelated Factor Demand Functions," *Amer. Econ. Rev.*, Sept. 1969, *59*, 457–71.

S. Rosen, "Short-run Employment Variations on Class-I Railroads in the U.S., 1947–63," *Econometrica*, Oct. 1968, *36*, 511–29.

M. Rothschild, "Changing Demand: Its Costs and Consequences." Paper presented at the December 1968 meetings of the Econometric Society.

——— and J. Stiglitz, "Increasing Risk: A Definition and Its Economic Consequences," Cowles Foundation Discussion Paper No. 275, 1969.

K. Smith, "The Effect of Uncertainty on Monopoly Price, Capital Stock and Utilization of Capital," *J. Econ. Theor.*, June 1969, *1*, 48–59.

U.S. Bureau of Labor Statistics, *Employer Expenditures for Selected Supplementary Compensation Practices, for Production and Related Workers, Manufacturing Industries 1962*, Bull. 1428, Washington 1965.

U.S. Congress, hearings before Subcommittee on Labor of the House of Representatives, *Overtime Pay Penalty Act of 1964*, 88th Cong., 2nd Sess., Washington 1964.

Complementarity and Stability of Equilibrium

By JAMES P. QUIRK*

The most important result of the literature concerned with the stability of the competitive equilibrium is the proof by Kenneth Arrow and Leonid Hurwicz (1958) and Arrow, H. D. Block and Hurwicz that, under a *tatonnement* mechanism, if all commodities are "gross substitutes," then the competitive equilibrium is globally stable. No other theorem of comparable generality has been proved in the stability literature. In this paper, we examine the stability properties of competitive equilibrium positions at which one or more pairs of commodities are "gross complements," that is, an increase in the price of good i (other prices being held fixed) leads to a decrease in excess demand for good j, and conversely. The main result of the paper may be formulated as follows.

Proposition: Assume that for any pair of commodities i and j, either i and j are gross substitutes or, i and j are gross complements or, i and j are independent goods. Then the presence of a pair of complementary commodities precludes the proof of stability of the competitive equilibrium under a *tatonnement* process from the substitutability-complementarity properties of the model alone.[1]

This proposition is related to certain results in the Arrow-Hurwicz paper, where it was shown that the case in which all commodities are gross complements cannot occur at a stable equilibrium position, nor can the "Morishima" case characterize a stable equilibrium (i.e., the case in which all goods obey the rules "substitutes of substitutes and complements of complements are substitutes while complements of substitutes and substitutes of complements are complements"). In contrast to the Arrow-Hurwicz results, here we do not prove *instability* but instead show that stability cannot be proved from the qualitative properties of the competitive model alone, under the assumptions given in the proposition, except in the gross substitute case. The proposition thus asserts that if stability is to be proved for a competitive equilibrium position at which complementarity is present, it is necessary to introduce restrictive quantitative assumptions into the analysis, such as, e.g., the assumption that all individuals are identical, that aggregate excess demand functions obey the weak axiom of revealed preference, that initial holdings represent a Pareto optimum, or some other such assumption as in the treatment by Arrow and Hurwicz. The gross substitute case (strictly speaking, its generalization, the "weak" gross substitute case) is thus the only qualitatively specified competitive system for which stability can be proved on qualitative grounds alone, given that complementarity-substitutability relations are "symmetric" as between commodities.

We first present an outline of the model of the competitive economy that will be

* Professor of economics, University of Kansas. This research was supported in part under a grant from the National Science Foundation.

[1] It should be noted that the proposition excludes the case of asymmetry, e.g., the case where for some i, j, i is a substitute for j but j is a complement (or independent) of i.

employed in this paper and following this, we establish the proposition relating to the possibility of proving stability of the competitive model when complementarity is present. The final part of the paper is concerned with summarizing certain classes of qualitatively specified cases in which stability results may be obtained in the context of the competitive model.

Following Arrow and Hurwicz (1958), the competitive model is assumed to consist of $n+1$ excess demand functions for the $n+1$ commodities of the system. Let p_i denote the price of the i^{th} commodity, $i=0, \ldots, n$ and let $p=(p_0, \ldots, p_n)$ denote a price vector. We denote excess demand for the i^{th} commodity by $E_i(p)$, $i=0, \ldots, n$. No attempt is made to relate properties of these aggregated excess demand functions to those of the underlying individual excess demand functions, but it is postulated that Walras' Law, (W), and homogeneity of excess demand functions, (H), hold:

$$(W) \qquad \sum_{i=0}^{n} p_i E_i(p) = 0 \quad \text{for every } p \geq 0$$

$$(H) \qquad E_i(\lambda p) = E_i(p) \quad \text{for every } \lambda > 0, \quad i = 0, \ldots, n$$

An equilibrium of the competitive system is defined as a price vector $\bar{p}=(\bar{p}_0, \ldots, \bar{p}_n)$ such that excess demand is zero in each market:

$$(1) \qquad E_i(\bar{p}) = 0 \quad i = 0, \ldots, n$$

Assume that excess demand functions have partial derivatives of all orders in a neighborhood of equilibrium and denote $\partial E_i/\partial p_j$ by E_{ij}. Then, in differential form, (W) and (H) can be written as:

$$(W') \qquad \sum_{i=0}^{n} E_{ij}(\bar{p})\bar{p}_i = 0 \quad j = 0, \ldots, n$$

$$(H') \qquad \sum_{j=0}^{n} E_{ij}(\bar{p})\bar{p}_j = 0 \quad i = 0, \ldots, n$$

(where the notation indicates that the partial derivatives are all evaluated at the equilibrium position).

Choose commodity 0 as numeraire, so that $p_0=1$. The dynamic behavior of the competitive model is assumed to be governed by a *tatonnement* process, with adjustment equations given by:

$$(2) \qquad Dp_i = g_i(E_i(p)) \quad i = 1, \ldots, n$$

where Dp_i denotes dp_i/dt and $g_i(\alpha)$ satisfies $g_i(0)=0$ and $dg_i/dE_i>0$.

Taking a linear approximation of (2) in a neighborhood of equilibrium gives

$$(3) \qquad Dp_i = \frac{dg_i}{dE_i} \sum_{j=1}^{n} E_{ij}(\bar{p})(p_j - \bar{p}_j) \qquad i = 1, \ldots, n.$$

Asymptotic stability of the linear approximation system (3) occurs for arbitrary initial positions if, and only if, every characteristic root of the matrix $[(dg_i/dE_i)E_{ij}]$ $i, j=1, \ldots, n$ has a negative real part.

We assume that the equilibrium price vector is strictly positive. We may then choose the units of measurement for commodities so that $\bar{p}_i=1$ $i=1, \ldots, n$. For simplicity of notation, let $a_{ij}=E_{ij}(\bar{p})$ $i, j=0, \ldots, n$ and let $A=[a_{ij}]$ $i, j=1, \ldots, n$, $A^*=[a_{ij}]$ $i, j=0, \ldots, n$. Then, under the choice of units of measurement, we can write (W'), (H') and (3) as

$$(W'') \qquad \sum_{i=0}^{n} a_{ij} = 0 \quad j = 0, \ldots, n$$

$$(H'') \qquad \sum_{j=0}^{n} a_{ij} = 0 \quad i = 0, \ldots, n$$

$$(3') \qquad Dp = BA(p - \bar{p})$$

where $Dp=(dp_i/dt)$ $i=1, \ldots, n$ $B=[dg_i/dE_i]$ is a diagonal matrix with all diagonal elements positive, A is as defined above, and $(p-\bar{p})=(p_i-\bar{p}_i)$ $i=1, \ldots, n$.

Then the restrictions (W'') and (H'')

specify that the matrix A^* has all row and column sums zero. We are particularly concerned, of course, with the properties of the submatrix A of A^*. For convenience, we say that A is a *stable matrix* if all the characteristic roots of A have negative real parts. We note that A stable does not in general imply that DA is stable for every diagonal matrix D with diagonal elements positive. When this does occur, we refer to A as a D-stable matrix.

The problem posed here is the following. Assume that the complementarity-substitutability properties of A^* are specified (i.e., we know the *signs* $(+, -, 0)$ of all the entries in A^*). In particular, i and j are (gross) substitutes if $a_{ij}>0$, $a_{ji}>0$, while i and j are (gross) complements if $a_{ij}<0$, $a_{ji}<0$. Suppose that all commodities are well defined (gross) substitutes or complements; i.e., $sgn\ a_{ij}=sgn\ a_{ji}$ for any $i\neq j$. Then under which sign patterns for A^* can we prove stability and/or D-stability for the matrix A in (3′)?

It might first be noted that the assumption $sgn\ a_{ij}=sgn\ a_{ji}\ i\neq j$, which we will refer to as "sign symmetry" of A^*, is a restrictive assumption so that the results stated in this section do not necessarily apply to the general case of qualitatively specified competitive models. (We return to this in the final section of the paper.) However, the case under consideration is of considerable interest in itself, being applicable except when individual income effects are "too" asymmetric, and also coinciding with the class of cases considered by J. R. Hicks. In fact, it is difficult to find treatments of the case where "sign asymmetry" occurs in the economics literature.

Because we will deal with classes of matrices specified in terms of sign pattern configurations, we introduce the notion of a *qualitative matrix* (see also Quirk 1967, 1968). Given a real $n\times n$ matrix C, let $sgn\ C=[sgn\ c_{ij}]$

$$\text{where } sgn\ c_{ij} = \begin{cases} -1 & \text{if } c_{ij} < 0 \\ +1 & \text{if } c_{ij} > 0 \\ 0 & \text{if } c_{ij} = 0 \end{cases}$$

Let $Q_C=[E|sgn\ E=sgn\ C]$. Then Q_C is an equivalence class of matrices all with the same sign pattern as the matrix C, and Q_C is referred to as a qualitative matrix.

The competitive system outlined above is said to be *qualitatively specified* once the signs $(+, -, 0)$ of all the entries in A^* are given. But Walras' Law and homogeneity impose additional quantitative restrictions on A^* (and A) through the conditions that row and column sums in A^* are all zero, and these quantitative restrictions must be exploited in determining the presence or absence of stability with respect to the matrix A.

As a matter of notation, we identify $n+1\times n+1$ matrices with asterisks, so that $C^*=[c_{ij}]\ \ i, j=0, \ldots, n$, while we use the symbol without asterisk to denote the submatrix in rows and columns $1, \ldots, n$; i.e., $C=[c_{ij}]\ i, j=1, \ldots, n$. Given A^*, let $S_{A^*}=\{C^*|C^*\in Q_{A^*}$ and C^* satisfies $\sum_{j=0}^{n} c_{ij}=0\ \ i=0, \ldots, n$ and $\sum_{i=0}^{n} c_{ij}=0\ \ j=0, \ldots, n\}$. Then A is said to be *qualitatively stable under* (W) *and* (H) if $C^*\in S_{A^*}$ $\Rightarrow C$ is a stable matrix. (Similarly, A is qualitatively D-stable under (W) and (H) if $C^*\in S_{A^*}\Rightarrow C$ is a D-stable matrix.)

The problem posed earlier thus reduces to that of determining the family of qualitative matrices Q_{A^*} such that A is qualitatively stable under (W) and (H) (or qualitatively D-stable under (W) and (H)) given that A^* is sign symmetric ($sgn\ a_{ij}=sgn\ a_{ji}$ for $i\neq j$).

First we note the relationship between qualitative stability of A and Hicksian stability of A^*.

Lemma 1. Assume A^* is sign symmetric. Then A is qualitatively stable under (W) and (H) only if $C^*\in S_{A^*}\Rightarrow C$ is quasi-

negative definite; i.e., $x'Cx<0$ for every $x\neq 0$.

Proof: If A is qualitatively stable under (W) and (H), then $C^*\in S_A^*\Rightarrow C$ is a stable matrix. But under the sign symmetry condition, $C^*\in S_A^*\Rightarrow C^*+C^{*\prime}\in S_A^*$ since $C^*+C^{*\prime}$ satisfies $\sum_{i=0}^n (c_{ij}+c_{ji})=0$ $j=0,\ldots n$, $\sum_{j=0}^n (c_{ij}+c_{ji})=0$ $i=0,\ldots,n$ while $sgn\ (c_{ij}+c_{ji})=sgn\ c_{ij}$. Hence $C+C'$ is a stable matrix; i.e., $x'(C+C')x<0$, so that C is quasi-negative definite.

Lemma 2. Assume C is quasi-negative definite. Then every $n\times n$ principal submatrix of C^* is quasi-negative definite, if row and column sums in C^* are zero.

Proof: Let $x^*=(x_0, x_1, \ldots, x_n)$. Consider the quadratic form $x^{*\prime}C^*x^*=$

$$\sum_{i=0}^n\sum_{j=0}^n c_{ij}x_ix_j=\sum_{i=1}^n\sum_{j=1}^n c_{ij}x_ix_j + x_0\left[\sum_{j=1}^n c_{0j}x_j+\sum_{i=1}^n c_{i0}x_i\right]+x_0^2c_{00}$$

By (W'') and (H'')

$$c_{0j}=-\sum_{i=1}^n c_{ij},\quad c_{i0}=-\sum_{j=1}^n c_{ij}$$

while

$$c_{00}=-\sum_{i=1}^n c_{i0}=-\sum_{j=1}^n c_{0j}=\sum_{i=1}^n\sum_{j=1}^n c_{ij}$$

Hence $x^{*\prime}C^*x^*=$

$$\sum_{i=1}^n\sum_{j=1}^n c_{ij}x_ix_j - x_0\left\{\sum_{i=1}^n\sum_{j=1}^n c_{ij}x_j+\sum_{i=1}^n\sum_{j=1}^n c_{ij}x_i\right\} + x_0^2\sum_{i=1}^n\sum_{j=1}^n c_{ij}$$

Note that if $x_0=0$, then by quasi-negative definiteness of C, the quadratic form is negative for $x^*\neq 0$. Hence assume $x_0\neq 0$ and let $y_i=(x_i/x_0)$ $i=1,\ldots,n$. Consider

$$\begin{aligned}\frac{1}{x_0^2}x^{*\prime}C^*x^* &= \sum_{i=1}^n\sum_{j=1}^n c_{ij}\{y_iy_j-y_i-y_j+1\}\\ &=\sum_{i=1}^n\sum_{j=1}^n c_{ij}(y_i(y_j-1)-(y_j-1))\\ &=\sum_{i=1}^n\sum_{j=1}^n c_{ij}(y_i-1)(y_j-1)\end{aligned}$$

Then by quasi-negative definiteness of C, we have $x^{*\prime}C^*x^*\leq 0$ and $x^{*\prime}C^*x^*<0$ unless $x_i=x_j$ for all i, $j=0,\ldots,n$. Since the quadratic form associated with any $n\times n$ principal submatrix of C^* is obtained by setting $x_i=0$ for the row (and column) index i deleted from C^*, it is clear that that quadratic form is negative except when evaluated at the zero vector, from which the lemma follows.

Corollary: Assume A^* is sign symmetric. Then A qualitatively stable under (W) and (H) and $C^*\in S_A^*\Rightarrow$ every $n\times n$ principal submatrix of C^* is Hicksian perfectly stable; i.e., every principal minor of C^* of order i has sign $(-1)^i$ $i=1,\ldots,n$.[2]

It follows directly that if complementarity is present, qualitative stability of A cannot be established under the sign symmetry condition. Assume that a pair of commodities i, j exhibit complementarity ($a_{ij}<0$, $a_{ji}<0$) and consider the 2×2 principal submatrix involving the i^{th} and j^{th} rows and columns. If A is qualitatively stable under (W) and (H), then $C^*\in S_A^*$ implies C is a stable matrix. Since qualitative stability implies $C+C'$ stable by Lemma 1, there is no loss of generality in assuming that C is symmetric. Let α_i, α_j denote the row (column) sums in the 2×2 submatrix. Then $c_{ii}+c_{ij}=\alpha_i$, $c_{ij}+c_{jj}=\alpha_j$ and, by assumption, $c_{ii}c_{jj}-c_{ij}^2>0$, while $c_{ii}<0$, $c_{jj}<0$ by the corollary above.

[2] If C^* is symmetric ($c_{ij}=c_{ji}$ $i\neq j$) then the principal minor condition is both necessary and sufficient for C to be (quasi) negative definite. Because we deal here only with the weaker condition of *sign* symmetry, only necessity follows from Lemmas 1 and 2.

Further, since $a_{ij}<0$, $a_{ji}<0$, it follows that $c_{ij}=c_{ji}<0$, so that $\alpha_i<0$, $\alpha_j<0$. Consider $B^* \in Q_{A^*}$ where all entries in B^* are identical to those in C^* except that $b_{ii}=\dfrac{\alpha_i^2}{\alpha_i+\alpha_j}+\epsilon$, $b_{ij}(=b_{ji})=\dfrac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}-\epsilon$, $b_{jj}=\dfrac{\alpha_j^2}{\alpha_i+\alpha_j}+\epsilon$, where $\epsilon>0$ is chosen sufficiently small so that $b_{ii}<0$, $b_{ji}=b_{ij}<0$, $b_{jj}<0$. Hence $b_{ii}+b_{ij}=\dfrac{\alpha_i^2+\alpha_i\alpha_j}{\alpha_i+\alpha_j}=\alpha_i$; $\quad b_{ij}+b_{jj}=\dfrac{\alpha_j^2+\alpha_i\alpha_j}{\alpha_i+\alpha_j}=\alpha_j$. Clearly $B^* \in S_{A^*}$.

Further, $b_{ii}b_{jj}-b_{ij}b_{ji}=\epsilon(\alpha_i+\alpha_j)<0$, so that B^* is not Hicksian and consequently A is not qualitatively stable under (W) and (H). Hence we have the following:

Theorem. Assume A^* is sign symmetric. If A^* contains any negative off diagonal elements, then A is not qualitatively stable under (W) and (H).

Corollary: Assume A^* is sign symmetric and indecomposable. Then A is qualitatively stable (qualitatively D-stable) under (W) and (H) if, and only if, A^* is a weak gross substitute matrix.[3]

The above analysis gives a complete characterization of qualitative stability of the competitive equilibrium in the "usual" case in which sign symmetry is present. However, because economic theory does not preclude, on a priori grounds, competitive systems in which (sign) asymmetries may occur, it would be of interest to determine the extent to which the above results can be extended to such systems. Complete results for such cases do not yet exist. However, qualitative stability under (W) and (H) has been proved in all of the following cases, where

$$sgn\ A^* = sgn\left[\begin{array}{c|ccc} a_{00} & a_{01} & a_{02} & a_{03} \\ \hline a_{10} & & & \\ a_{20} & & A & \\ a_{30} & & & \end{array}\right]$$

$$(1)\quad sgn\ A^* = \left[\begin{array}{c|ccc} - & ? & ? & ? \\ \hline + & - & + & + \\ + & + & - & + \\ + & + & + & - \end{array}\right]$$

$$(2)\quad sgn\ A^* = \left[\begin{array}{c|ccc} - & + & + & + \\ \hline ? & - & + & + \\ ? & + & - & + \\ ? & + & + & - \end{array}\right]$$

$$(3)\quad sgn\ A^* = \left[\begin{array}{c|ccc} - & + & + & - \\ \hline - & - & + & + \\ - & + & - & + \\ + & + & + & - \end{array}\right]$$

$$(4)\quad sgn\ A^* = \left[\begin{array}{c|ccc} - & ? & ? & + \\ \hline + & - & - & 0 \\ ? & + & - & - \\ ? & 0 & + & - \end{array}\right]$$

(where ? denotes an element of arbitrary sign).

Cases (1), (2), and (3) represent variants of the gross substitute case. Since the dominant diagonal property of the gross substitute case can be proved using either row sums or column sums, the numeraire row (or column) can be characterized by elements of arbitrary sign so long as the column (or row) has positive off diagonal

[3] D-stability follows from Lemma 1, since any quasi-negative definite matrix is D-stable. Sufficiency of the weak gross substitute conditions is proved by Arrow and Hurwicz (1960).

entries, as in cases (1) and (2). Case (3) illustrates the fact that if strict asymmetry of signs characterizes the numeraire row-column relationships and if A is a gross substitute matrix, then qualitative stability can be proved if, and only if, exactly one off diagonal entry in the numeraire row (or column) is negative. If A^* has no zero entries and if A is gross substitute, then cases (1), (2), and (3), are the *only* cases in which qualitative stability under (W) and (H) can be proved (see Quirk 1967). It might also be noted that if the numeraire is a gross substitute for all other commodities (and conversely), if there are no zeros in A^* and if there are more than three commodities (including numeraire), then qualitative stability can be proved only if A^* is a gross substitute matrix (again, see Quirk 1967).

Case (4) illustrates the "sign-stable" assumptions as given in Quirk and R. Rupert; any matrix with the same sign pattern as A is stable, whatever are the row and column sum restrictions on A, hence A is qualitatively stable under (W) and (H). We note that one sign pattern for A^* (from case (4)) consistent with qualitative stability of A is the following:

$$sgn\ A^* = \begin{bmatrix} - & + & - & + \\ + & - & - & 0 \\ - & + & - & - \\ + & 0 & + & - \end{bmatrix}$$

In this example, the numeraire is a well defined complement with respect to good #2 at a stable equilibrium, so that the proposition of this paper does not extend to systems in which asymmetry in signs is present. It is conjectured, however, that qualitative stability cannot be proved if A (rather than A^*) contains a pair of complementary commodities, but work remains to establish this conjecture.

REFERENCES

K. Arrow, H. D. Block and L. Hurwicz, "On the Stability of the Competitive Equilibrium, II," *Econometrica*, Jan. 1959, *27*, 82–109.

K. Arrow and L. Hurwicz, "On the Stability of the Competitive Equilibrium, I," *Econometrica*, Oct. 1958, *26*, 522–52.

——— and ———, "Competitive Stability under Weak Gross Substitutability: The 'Euclidean Distance' Approach," *Int. Econ. Rev.*, Jan. 1960, *1*, 38–49.

L. Bassett, J. Maybee and J. Quirk, "Qualitative Economics and the Scope of the Correspondence Principle," *Econometrica*, July–Oct. 1968, *36*, 544–63.

J. R. Hicks, *Value and Capital*, 2d ed., London 1946.

J. Quirk, "Comparative Statics under Walras' Law: The Case of Strong Dependence," *Rev. Econ. Stud.*, Jan. 1968, *35*, 11–21.

———, "The Competitive Equilibrium: A Qualitative Analysis," mimeo. Univ. Kansas, 1967.

———and R. Ruppert, "Qualitative Economics and the Stability of Equilibrium," *Rev. Econ. Stud.*, Oct. 1965, *32*, 311–26.

Uncertainty and the Evaluation of Public Investment Decisions

By KENNETH J. ARROW AND ROBERT C. LIND*

The implications of uncertainty for public investment decisions remain controversial. The essence of the controversy is as follows. It is widely accepted that individuals are not indifferent to uncertainty and will not, in general, value assets with uncertain returns at their expected values. Depending upon an individual's initial asset holdings and utility function, he will value an asset at more or less than its expected value. Therefore, in private capital markets, investors do not choose investments to maximize the present value of expected returns, but to maximize the present value of returns properly adjusted for risk. The issue is whether it is appropriate to discount public investments in the same way as private investments.

There are several positions on this issue. The first is that risk should be discounted in the same way for public investments as it is for private investments. It is argued that to treat risk differently in the public sector will result in overinvestment in this sector at the expense of private investments yielding higher returns. The leading proponent of this point of view is Jack Hirshleifer.[1] He argues that in perfect capital markets, investments are discounted with respect to both time and risk and that the discount rates obtaining in these markets should be used to evaluate public investment opportunities.

A second position is that the government can better cope with uncertainty than private investors and, therefore, government investments should not be evaluated by the same criterion used in private markets. More specifically, it is argued that the government should ignore uncertainty and behave as if indifferent to risk. The government should then evaluate investment opportunities according to their present value computed by discounting the expected value of net returns, using a rate of discount equal to the private rate appropriate for investments with certain returns. In support of this position it is argued that the government invests in a greater number of diverse projects and is able to pool risks to a much greater extent than private investors.[2] Another supporting line of argument is that many of the uncertainties which arise in private capital markets are related to what may be termed moral hazards. Individuals involved in a given transaction may hedge against the possibility of fraudulent behavior on the part of their associates. Many such risks are not present in the case of public investments and, therefore, it can be argued that it is not appropriate for the government to take these risks into account when choosing among public investments.

There is, in addition, a third position on the government's response to uncertainty. This position rejects the notion that indi-

* The authors are, respectively, professor of economics at Harvard University; and assistant professor of engineering-economic systems and, by courtesy, of economics at Stanford University. R. C. Lind's work has been supported by National Science Foundation grant NSF-GK-1683.

[1] J. Hirshleifer (1965, 1966) and Hirshleifer, J. C. De Haven, and J. W. Milliman (pp. 139–50).

[2] For this point of view, see P. A. Samuelson and W. Vickrey.

vidual preferences as revealed by market behavior are of normative significance for government investment decisions, and asserts that time and risk preferences relevant for government action should be established as a matter of national policy. In this case the correct rules for action would be those established by the appropriate authorities in accordance with their concept of national policy. The rate of discount and attitude toward risk would be specified by the appropriate authorities and the procedures for evaluation would incorporate these time and risk preferences. Two alternative lines of argument lead to this position. First, if one accepts the proposition that the state is more than a collection of individuals and has an existence and interests apart from those of its individual members, then it follows that government policy need not reflect individual preferences. A second position is that markets are so imperfect that the behavior observed in these markets yields no relevant information about the time and risk preferences of individuals. It follows that some policy as to time and risk preference must be established in accordance with other evidence of social objectives. One such procedure would be to set national objectives concerning the desired rate of growth and to infer from this the appropriate rate of discount.[3] If this rate were applied to the expected returns from all alternative investments, the government would in effect be behaving as if indifferent to risk.

The approach taken in this paper closely parallels the approach taken by Hirshleifer, although the results differ from his. By using the state-preference approach to market behavior under uncertainty, Hirshleifer demonstrates that investments will not, in general, be valued at the sum of the expected returns discounted at a rate appropriate for investments with certain returns.[4] He then demonstrates that using this discount rate for public investments may lead to non-optimal results, for two reasons. First, pooling itself may not be desirable.[5] If the government has the opportunity to undertake only investments which pay off in states where the payoff is highly valued, to combine such investments with ones that pay off in other states may reduce the value of the total investment package. Hirshleifer argues that where investments can be undertaken separately they should be evaluated separately, and that returns should be discounted at rates determined in the market. Second, even if pooling were possible and desirable, Hirshleifer argues correctly that the use of a rate of discount for the public sector which is lower than rates in the private sector can lead to the displacement of private investments by public investments yielding lower expected returns.[6]

For the case where government pooling is effective and desirable, he argues that rather than evaluate public investments differently from private ones, the government should subsidize the more productive private investments. From this it follows that to treat risk differently for public as opposed to private investments would only be justified if it were impossible to transfer the advantages of government pooling to private investors. Therefore, at most, the argument for treating public risks differently than private ones in evaluating investments is an argument for the "second best."[7]

The first section of this paper addresses the problem of uncertainty, using the state-preference approach to market behavior. It demonstrates that if the returns

[3] For this point of view, see O. Eckstein and S. Marglin.

[4] Hirshleifer (1965 pp. 523–34); (1966, pp. 268–75).

[5] Hirshleifer (1966, pp. 270–75).

[6] Hirshleifer (1966, pp. 270–75).

[7] Hirshleifer (1966, p. 270).

from any particular investment are independent of other components of national income, then the present value of this investment equals the sum of expected returns discounted by a rate appropriate for investments yielding certain returns. This result holds for both private and public investments. Therefore, by adding one plausible assumption to Hirshleifer's formulation, the conclusion can be drawn that the government should behave as an expected-value decision maker and use a discount rate appropriate for investments with certain returns. This conclusion needs to be appropriately modified when one considers the case where there is a corporate income tax.

While this result is of theoretical interest, as a policy recommendation it suffers from a defect common to the conclusions drawn by Hirshleifer. The model of the economy upon which these recommendations are based presupposes the existence of perfect markets for claims contingent on states of the world. Put differently, it is assumed that there are perfect insurance markets through which individuals may individually pool risks. Given such markets, the distribution of risks among individuals will be Pareto optimal. The difficulty is that many of these markets for insurance do not exist, so even if the markets which do exist are perfect, the resulting equilibrium will be sub-optimal. In addition, given the strong evidence that the existing capital markets are not perfect, it is unlikely that the pattern of investment will be Pareto optimal. At the margin, different individuals will have different rates of time and risk preference, depending on their opportunities to borrow or to invest, including their opportunities to insure.

There are two reasons why markets for many types of insurance do not exist. The first is the existence of certain moral hazards.[8] In particular, the fact that someone has insurance may alter his behavior so that the observed outcome is adverse to the insurer. The second is that such markets would require complicated and specialized contracts which are costly. It may be that the cost of insuring in some cases is so high that individuals choose to bear risks rather than pay the transaction costs associated with insurance.

Given the absence of some markets for insurance and the resulting sub-optimal allocation of risks, the question remains: How should the government treat uncertainty in evaluating public investment decisions? The approach taken in this paper is that individual preferences are relevant for public investment decisions, and government decisions should reflect individual valuations of costs and benefits. It is demonstrated in the second section of this paper that when the risks associated with a public investment are publicly borne, the total cost of risk-bearing is insignificant and, therefore, the government should ignore uncertainty in evaluating public investments. Similarly, the choice of the rate of discount should in this case be independent of considerations of risk. This result is obtained not because the government is able to pool investments but because the government distributes the risk associated with any investment among a large number of people. It is the risk-spreading aspect of government investment that is essential to this result.

There remains the problem that private investments may be displaced by public ones yielding a lower return if this rule is followed, although given the absence of insurance markets this will represent a Hicks-Kaldor improvement over the initial situation. Again the question must be

[8] For a discussion of this problem see M. V. Pauly and Arrow (1968).

asked whether the superior position of the government with respect to risk can be made to serve private investors. This leads to a discussion of the government's role as a supplier of insurance, and of Hirshleifer's recommendation that private investment be subsidized in some cases.

Finally, the results obtained above apply to risks actually borne by the government. Many of the risks associated with public investments are borne by private individuals, and in such cases it is appropriate to discount for risk as would these individuals. This problem is discussed in the final section of the paper. In addition, a method of evaluating public investment decisions is developed that calls for different rates of discount applied to different classes of benefits and costs.

I. *Markets for Contingent Claims and Time-Risk Preference*[9]

For simplicity, consider an economy where there is one commodity and there are I individuals, S possible states of the world, and time is divided into Q periods of equal length. Further suppose that each individual acts on the basis of his subjective probability as to the states of nature; let π_{is} denote the subjective probability assigned to state s by individual i. Now suppose that each individual in the absence of trading owns claims for varying amounts of the one commodity at different points in time, given different states of the world. Let $\bar{x}_{isq}$ denote the initial claim to the commodity in period q+1 if state s occurs which is owned by individual i. Suppose further that all trading in these claims takes place at the beginning of the first period, and claims are bought and sold on dated commodity units contingent on a state of the world. All claims can be constructed from basic claims which pay one commodity unit in period q+1, given state s, and nothing in other states or at other times; there will be a corresponding price for this claim, $p_{sq}(s=1, \ldots, S;\ q=0, \ldots, Q-1)$. After the trading, the individual will own claims x_{isq}, which he will exercise when the time comes to provide for his consumption. Let $V_i(x_{i1,0}, \ldots, x_{i1,Q-1}, x_{i2,0}, \ldots, x_{iS,Q-1})$ be the utility of individual i if he receives claims x_{isq} $(s=1, \ldots, S;\ q=0, \ldots, Q-1)$. The standard assumptions are made that V_i is strictly quasi-concave $(i=1, \ldots, I)$.

Therefore each individual will attempt to maximize,

$$(1)\quad V_i(x_{i1,0}, \ldots, x_{i1,Q-1}, x_{i2,0}, \ldots, x_{iS,Q-1})$$

subject to the constraint

$$\sum_{q=0}^{Q-1} \sum_{s=1}^{S} p_{sq} x_{isq} = \sum_{q=0}^{Q-1} \sum_{s=1}^{S} p_{sq} \bar{x}_{isq}$$

Using the von Neumann-Morgenstern theorem and an extension by Hirshleifer,[10] functions U_{is} $(s=1, \ldots, S)$ can be found such that

$$(2)\quad \begin{aligned} &V_i(x_{i1,0}, \ldots, x_{iS,Q-1}) \\ &\quad = \sum_{s=1}^{S} \pi_{is} U_{is}(x_{is0}, x_{is1}, \ldots, x_{iS,Q-1}) \end{aligned}$$

In equation (2) an individual's utility, given any state of the world, is a function of his consumption at each point in time. The subscript s attached to the function U_{is} is in recognition of the fact that the value of a given stream of consumption may depend on the state of the world.

The conditions for equilibrium require that

$$(3)\quad \pi_{is} \frac{\partial U_{is}}{\partial x_{isq}} = \lambda_i p_{sq} \qquad (i = 1, \ldots, I;\ s = 1, \ldots, S;\ q = 0, \ldots, Q-1)$$

[9] For a basic statement of the state-preference approach, see Arrow (1964) and G. Debreu.

[10] J. von Neumann and O. Morgenstern, and Hirshleifer (1965, pp. 534–36).

where λ_i is a Lagrangian multiplier.

From (3) it follows that

$$(4)\quad \frac{p_{sq}}{p_{rm}} = \frac{\pi_{is}\dfrac{\partial U_{is}}{\partial x_{isq}}}{\pi_{rm}\dfrac{\partial U_{ir}}{\partial x_{irm}}} \quad (i = 1, \ldots, I;$$

$$r, s = 1, \ldots, S; m, q = 0, \ldots, Q-1)$$

Insight can be gained by analyzing the meaning of the prices in such an economy. Since trading takes place at time zero, p_{sq} represents the present value of a claim to one commodity unit at time q, given state s. Clearly,

$$\sum_{s=1}^{S} p_{s0} = 1$$

since someone holding one commodity unit at time zero has a claim on one commodity unit, given any state of the world. It follows that p_{sq} is the present value of one commodity at time q, given state s, in terms of a certain claim on one commodity unit at time zero. Therefore, the implicit rate of discount to time zero on returns at time q, given state s, is defined by $p_{sq} = 1/1+r_{sq}$.

Now suppose one considers a certain claim to one commodity unit at time q; clearly, its value is

$$p_q = \sum_{s=1}^{S} p_{sq}$$

and the rate of discount appropriate for a certain return at time q is defined by

$$(5)\quad \frac{1}{1+r_q} = \sum_{s=1}^{S} \frac{1}{1+r_{sq}} = \sum_{s=1}^{S} p_{sq}$$

Given these observations, we can now analyze the appropriate procedure for evaluating government investments where there are perfect markets for claims contingent on states of the world.[11] Consider an investment where the overall effect on market prices can be assumed to be negligible, and suppose the net return from this investment for a given time and state is h_{sq} $(s=1, \ldots, S; q=0, \ldots, Q-1)$. Then the investment should be undertaken if

$$(6)\quad \sum_{q=0}^{Q-1} \sum_{s=1}^{S} h_{sq} p_{sq} > 0,$$

and the sum on the left is an exact expression for the present value of the investment. Expressed differently, the investment should be adopted if

$$(7)\quad \sum_{q=0}^{Q-1} \sum_{s=1}^{S} \frac{h_{sq}}{1+r_{sq}} > 0$$

The payoff in each time-state is discounted by the associated rate of discount. This is the essential result upon which Hirshleifer bases his policy conclusions.[12]

Now suppose that the net returns of the investment were (a) independent of the returns from previous investment, (b) independent of the individual utility functions, and (c) had an objective probability distribution, i.e., one agreed upon by everyone. More specifically, we assume that the set of all possible states of the world can be partitioned into a class of mutually exclusive and collectively exhaustive sets, E_t, indexed by the subscript t such that, for all s in any given E_t, all utility functions U_{is} are the same for any individual i $(i=1, \cdots, I)$, and such that all production conditions are the same. Put differently, for all s in E_t, U_{is} is the same for a given individual, but not necessarily for all individuals. At the same time there is another partition of the states of the world into sets, F_u, such that the return, h_{sq}, is the same for all s in F_u. Finally, we assume that the probability distribution of F_u is independent of E_t and is the same for all individuals.

Let E_{tu} be the set of all states of the world which lie in both E_t and F_u. For any given t and u, all states of the world in

[11] The following argument was sketched in Arrow (1966, pp. 28–30).

[12] Hirshleifer (1965, pp. 523–34).

E_{tu} are indistinguishable for all purposes, so we may regard it as containing a single state. Equations (3) and (5) and the intervening discussion still hold if we then replace s everywhere by tu. However, $U_{is} = U_{itu}$ actually depends only on the subscript, t, and can be written U_{it}. From the assumptions it is obvious and can be proved rigorously that the allocation x_{itq} also depends only on t, i.e., is the same for all states in E_t for any given t, so it may be written x_{itq}. Finally, let π_{it} be the probability of E_t according to individual i, and let π_u be the probability of F_u, assumed the same for all individuals. Then the assumption of statistical independence is written:

$$\pi_{itu} = \pi_{it}\pi_u \tag{8}$$

Then (3) can be written

$$\pi_{it}\pi_u \frac{\partial U_{it}}{\partial x_{itq}} = \lambda_i p_{tuq} \tag{9}$$

Since p_{tuq} and π_u are independent of i, so must be

$$\left(\pi_{it} \frac{\partial U_{it}}{\partial x_{itq}}\right) \Big/ \lambda_i;$$

on the other hand, this expression is also independent of u and so can be written μ_{tq}. Therefore,

$$p_{tuq} = \mu_{tq}\pi_u \tag{10}$$

Since the new investment has the same return for all states s in F_u, the returns can be written h_{uq}. Then the left-hand side of (6) can, with the aid of (10), be written

$$\begin{aligned} &\sum_{Q=0}^{Q-1} \sum_{s=1}^{S} h_{sq} p_{sq} \\ &= \sum_{q=0}^{Q-1} \sum_t \sum_u h_{uq} p_{tuq} \\ &= \sum_{q=0}^{Q-1} \left(\sum_t \mu_{tq}\right) \sum_u \pi_u h_{uq} \end{aligned} \tag{11}$$

But from (10)

$$\begin{aligned} p_q &= \sum_{s=1}^{S} p_{sq} = \sum_t \sum_u p_{tuq} \\ &= \left(\sum_t \mu_{tq}\right)\left(\sum_u \pi_u\right) = \sum_t \mu_{tq}, \end{aligned} \tag{12}$$

since of course the sum of the probabilities of the F_u's must be 1. From (11),

$$\sum_{q=0}^{Q-1} \sum_{s=1}^{S} h_{sq} p_{sq} = \sum_{Q=0}^{Q-1} \frac{1}{1 + r_q} \sum_u \pi_u h_{uq} \tag{13}$$

Equation (13) gives the rather startling result that the present value of any investment which meets the independence and objectivity conditions, equals the expected value of returns in each time period, discounted by the factor appropriate for a certain return at that time. This is true even though individuals may have had different probabilities for the events that governed the returns on earlier investments. It is also interesting to note that each individual will behave in this manner so that there will be no discrepancy between public and private procedures for choosing among investments.

The independence assumption applied to utility functions was required because the functions U_{is} are conditional on the states of the world. This assumption appears reasonable, and in the case where U_{is} is the same for all values of s, it is automatically satisfied. Then the independence condition is simply that the net returns from an investment be independent of the returns from previous investments.

The difficulty that arises if one bases policy conclusions on these results is that some markets do not exist, and individuals do not value assets at the expected value of returns discounted by a factor appropriate for certain returns. It is tempting to argue that while individuals do not behave as expected-value decision makers because of the nonexistence of certain markets for insurance, there is no reason why the government's behavior should not be consistent with the results derived above

where the allocation of resources was Pareto optimal. There are two difficulties with this line of argument. First, if we are to measure benefits and costs in terms of individuals' willingness to pay, then we must treat risk in accordance with these individual valuations. Since individuals do not have the opportunities for insuring assumed in the state-preference model, they will not value uncertainty as they would if these markets did exist. Second, the theory of the second best demonstrates that if resources are not allocated in a Pareto optimal manner, the appropriate public policies may not be those consistent with Pareto efficiency in perfect markets. Therefore, some other approach must be found for ascertaining the appropriate government policy toward risk. In particular, such an approach must be valid, given the nonexistence of certain markets for insurance and imperfections in existing markets.

II. *The Public Cost of Risk-Bearing*

The critical question is: What is the cost of uncertainty in terms of costs to individuals? If one adopts the position that costs and benefits should be computed on the basis of individual willingness to pay, consistency demands that the public costs of risk-bearing be computed in this way too. This is the approach taken here.

In the discussion that follows it is assumed that an individual's utility is dependent only upon his consumption and not upon the state of nature in which that consumption takes place. This assumption simplifies the presentation of the major theorem, but it is not essential. Again the expected utility theorem is assumed to hold. The presentation to follow analyzes the cost of risk-bearing by comparing the expected value of returns with the certainty equivalent of these returns. In this way the analysis of time and risk preference can be separated, so we need only consider one time period.

Suppose that the government were to undertake an investment with a certain outcome; then the benefits and costs are measured in terms of willingness to pay for this outcome. If, however, the outcome is uncertain, then the benefits and costs actually realized depend on which outcome in fact occurs. If an individual is risk-averse, he will value the investment with the uncertain outcome at less than the expected value of its net return (benefit minus cost) to him. Therefore, in general the expected value of net benefits overstates willingness to pay by an amount equal to the cost of risk-bearing. It is clear that the social cost of risk-bearing will depend both upon which individuals receive the benefits and pay the costs and upon how large is each individual's share of these benefits and costs.

As a first step, suppose that the government were to undertake an investment and capture all benefits and pay all costs, i.e., the beneficiaries pay to the government an amount equal to the benefits received and the government pays all costs. Individuals who incur costs and those who receive benefits are therefore left indifferent to their pre-investment state. This assumption simply transfers all benefits and costs to the government, and the outcome of the investment will affect government disbursements and receipts. Given that the general taxpayer finances government expenditures, a public investment can be considered an investment in which each individual taxpayer has a very small share.

For precision, suppose that the government undertook an investment and that returns accrue to the government as previously described. In addition, suppose that in a given year the government were to have a balanced budget (or a planned deficit or surplus) and that taxes would be reduced by the amount of the net benefits if the returns are positive, and raised if returns are negative. Therefore, when the government undertakes an investment,

each taxpayer has a small share of that investment with the returns being paid through changes in the level of taxes. By undertaking an investment the government adds to each individual's disposable income a random variable which is some fraction of the random variable representing the total net returns. The expected return to all taxpayers as a group equals expected net benefits.

Each taxpayer holds a small share of an asset with a random payoff, and the value of this asset to the individual is less than its expected return, assuming risk aversion. Stated differently, there is a cost of risk-bearing that must be subtracted from the expected return in order to compute the value of the investment to the individual taxpayer. Since each taxpayer will bear some of the cost of the risk associated with the investment, these costs must be summed over all taxpayers in order to arrive at the total cost of risk-bearing associated with a particular investment. These costs must be subtracted from the value of expected net benefits in order to obtain the correct measure for net benefits. The task is to assess these costs.

Suppose, as in the previous section, that there is one commodity, and that each individual's utility in a given year is a function of his income defined in terms of this commodity and is given by $U(Y)$. Further, suppose that U is bounded, continuous, strictly increasing, and differentiable. The assumptions that U is continuous and strictly increasing imply that U has a right and left derivative at every point and this is sufficient to prove the desired results; differentiability is assumed only to simplify presentation. Further suppose that U satisfies the conditions of the expected utility theorem.

Consider, for the moment, the case where all individuals are identical in that they have the same preferences, and their disposable incomes are identically distributed random variables represented by A. Suppose that the government were to undertake an investment with returns represented by B, which are statistically independent of A. Now divide the effect of this investment into two parts: a certain part equal to expected returns and a random part, with mean zero, which incorporates risk. Let $\bar{B}=E[B]$, and define the random variable X by $X=B-\bar{B}$. Clearly, X is independent of A and $E[X]=0$. The effect of this investment is to add an amount $\bar{B}$ to government receipts along with a random component represented by X. The income of each taxpayer will be affected through taxes and it is the level of these taxes that determines the fraction of the investment he effectively holds.

Consider a specific taxpayer and denote his fraction of this investment by s, $0\leqq s\leqq 1$. This individual's disposable income, given the public investment, is equal to $A+sB=A+s\bar{B}+sX$. The addition of sB to his disposable income is valued by the individual at its expected value less the cost of bearing the risk associated with the random component sX. If we suppose that each taxpayer has the same tax rate and that there are n taxpayers, then $s=1/n$, and the value of the investment taken over all individuals is simply $\bar{B}$ minus n times the cost of risk-bearing associated with the random variable $(1/n)X$. The central result of this section of the paper is that this total of the costs of risk-bearing goes to zero as n becomes large. Therefore, for large values of n the value of a public investment almost equals the expected value of that investment.

To demonstrate this, we introduce the function

$$W(s) = E[U(A + s\bar{B} + sX], \qquad 0 \leqq s \leqq 1 \tag{14}$$

In other words, given the random variables A and B representing his individual income before the investment and the income from the investment, respectively, his expected

utility is a function of s which represents his share of B. From (14) and the assumption that U' exists, it follows that

$$W'(s) = E[U'(A+s\bar{B}+sX)(\bar{B}+X)] \tag{15}$$

Since X is independent of A, it follows that $U'(A)$ and X are independent; therefore,

$$E[U'(A)X] = E[U'(A)]E[X] = 0$$

so that

$$\begin{aligned} W'(0) &= E[U'(A)(\bar{B}+X)] \\ &= \bar{B}E[U'(A)] \end{aligned} \tag{16}$$

Equation (16) is equivalent to the statement

$$\lim_{s\to 0} \frac{E[U(A+s\bar{B}+sX) - U(A)]}{s} = \bar{B}E[U'(A)] \tag{17}$$

Now let $s=1/n$, so that equation (17) becomes

$$\lim_{n\to\infty} nE\left[U\left(A+\frac{\bar{B}+X}{n}\right) - U(A)\right] = \bar{B}E[U'(A)] \tag{18}$$

If we assume that an individual whose preferences are represented by U is a risk-averter, then it is easily shown that there exists a unique number, $k(n)>0$, for each value of n such that

$$E\left[U\left(A+\frac{\bar{B}+X}{n}\right)\right] = E\left[U\left(A+\frac{\bar{B}}{n}-k(n)\right)\right], \tag{19}$$

or, in other words, an individual would be indifferent between paying an amount equal to $k(n)$ and accepting the risk represented by $(1/n)X$. Therefore, $k(n)$ can be said to be the cost of risk-bearing associated with the asset B. It can easily be demonstrated that $\lim_{n\to\infty} k(n)=0$, i.e., the cost of holding the risky asset goes to zero as the amount of this asset held by the individual goes to zero. It should be noted that the assumption of risk aversion is not essential to the argument but simply one of convenience. If U represented the utility function of a risk preferrer, then all the above statements would hold except $k(n)<0$, i.e., an individual would be indifferent between being paid $-k(n)$ and accepting the risk $(1/n)X$ (net of the benefit $(1/n)\bar{B}$).

We wish to prove not merely that the risk-premium of the representative individual, $k(n)$, vanishes, but more strongly that the total of the risk-premiums for all individuals, $nk(n)$, approaches zero as n becomes large.

From (18) and (19) it follows that

$$\lim_{n\to\infty} nE\left[U\left(A+\frac{\bar{B}}{n}-k(n)\right) - U(A)\right] = \bar{B}E[U'(A)] \tag{20}$$

In addition, $\bar{B}/n-k(n)\to 0$, when $n\to\infty$. It follows from the definition of a derivative that

$$\lim_{n\to\infty} \frac{E\left[U\left(A+\frac{\bar{B}}{n}-k(n)\right) - U(A)\right]}{\frac{\bar{B}}{n}-k(n)} = E[U'(A)] > 0 \tag{21}$$

Dividing (20) by (21) yields

$$\lim_{n\to\infty} [\bar{B} - nk(n)] = \bar{B} \tag{22}$$

or

$$\lim_{n\to\infty} nk(n) = 0 \tag{23}$$

The argument in (21) implies that $\bar{B}/n-k(n)\neq 0$. Suppose instead the equality held for infinitely many n. Substitution into the left-hand side of (20) shows that $\bar{B}$ must equal zero, so that $k(n)=0$ for all such n, and hence $nk(n)=0$ on that sequence, confirming (23).

Equation (23) states that the total of

the costs of risk-bearing goes to zero as the population of taxpayers becomes large. At the same time the monetary value of the investment to each taxpayer, neglecting the cost of risk, is $(1/n)\bar{B}$, and the total, summed over all individuals, is $\bar{B}$, the expected value of net benefits. Therefore, if n is large, the expected value of net benefits closely approximates the correct measure of net benefits defined in terms of willingness to pay for an asset with an uncertain return.

In the preceding analysis, it was assumed that all taxpayers were identical in that they had the same utility function, their incomes were represented by identically distributed variables, and they were subject to the same tax rates. These assumptions greatly simplify the presentation; however, they are not essential to the argument. Different individuals may have different preferences, incomes, and tax rates; and the basic theorem still holds, provided that as n becomes larger the share of the public investment borne by any individual becomes arbitrarily smaller.

The question necessarily arises as to how large n must be to justify proceeding as if the cost of publicly-borne risk is negligible. This question can be given no precise answer; however, there are circumstances under which it appears likely that the cost of risk-bearing will be small. If the size of the share borne by each taxpayer is a negligible component of his income, the cost of risk-bearing associated with holding it will be small. It appears reasonable to assume, under these conditions, that the total cost of risk-bearing is also small. This situation will exist where the investment is small with respect to the total wealth of the taxpayers. In the case of a federally sponsored investment, n is not only large but the investment is generally a very small fraction of national income even though the investment itself may be large in some absolute sense.

The results derived here and in the previous section depend on returns from a given public investment being independent of other components of national income. The government undertakes a wide range of public investments and it appears reasonable to assume that their returns are independent. Clearly, there are some government investments which are interdependent; however, where investments are interrelated they should be evaluated as a package. Even after such groupings are established, there will be a large number of essentially independent projects. It is sometimes argued that the returns from public investments are highly correlated with other components of national income through the business cycle. However, if we assume that stabilization policies are successful, then this difficulty does not arise. It should be noted that in most benefit-cost studies it is assumed that full employment will be maintained so that market prices can be used to measure benefits and costs. Consistency requires that this assumption be retained when considering risk as well. Further, if there is some positive correlation between the returns of an investment and other components of national income, the question remains as to whether this correlation is so high as to invalidate the previous result.

The main result is more general than the specific application to public investments. It has been demonstrated that if an individual or group holds an asset which is statistically independent of other assets, and if there is one or more individuals who do not share ownership, then the existing situation is not Pareto-efficient. By selling some share of the asset to one of the individuals not originally possessing a share, the cost of risk-bearing can be reduced while the expected returns remain unchanged. The reduction in the cost of risk-bearing can then be redistributed to bring about a Pareto improvement. This result is

similar to a result derived by Karl Borch. He proved that a condition for Pareto optimality in reinsurance markets requires that every individual hold a share of every independent risk.

When the government undertakes an investment it, in effect, spreads the risk among all taxpayers. Even if one were to accept that the initial distribution of risk was Pareto-efficient, the new distribution of risk will not be efficient as the government does not discriminate among the taxpayers according to their risk preferences. What has been shown is that in the limit the situation where the risk of the investment is spread over all taxpayers is such that there is only a small deviation from optimality with regard to the distribution of that particular risk. The overall distribution of risk may be sub-optimal because of market imperfections and the absence of certain insurance markets. The great advantage of the results of this section is that they are not dependent on the existence of perfect markets for contingent claims.

This leads to an example which runs counter to the policy conclusions generally offered by economists. Suppose that an individual in the private sector of the economy were to undertake a given investment and, calculated on the basis of expected returns, the investment had a rate of return of 10 per cent. Because of the absence of perfect insurance markets, the investor subtracted from the expected return in each period a risk premium and, on the basis of returns adjusted for risk, his rate of return is 5 percent. Now suppose that the government could invest the same amount of money in an investment which, on the basis of expected returns, would yield 6 percent. Since the risk would be spread over all taxpayers, the cost of risk-bearing would be negligible, and the true rate of return would be 6 percent. Further, suppose that if the public investment were adopted it would displace the private investment. The question is: Should the public investment be undertaken? On the basis of the previous analysis, the answer is yes. The private investor is indifferent between the investment with the expected return of 10 percent, and certain rate of return of 5 percent. When the public investment is undertaken, it is equivalent to an investment with a certain rate of return of 6 percent. Therefore, by undertaking the public investment, the government could more than pay the opportunity cost to the private investor of 5 percent associated with the diversion of funds from private investment.

The previous example illustrates Hirshleifer's point that the case for evaluating public investments differently from private ones is an argument for the second best. Clearly, if the advantages of the more efficient distribution of risk could be achieved in connection with the private investment alternative, this would be superior to the public investment. The question then arises as to how the government can provide insurance for private investors and thereby transfer the risks from the private sector to the public at large. The same difficulties arise as before, moral hazards and transaction costs. It may not be possible for the government to provide such insurance, and in such cases second-best solutions are in order. Note that if the government could undertake any investment, then this difficulty would not arise. Perhaps one of the strongest criticisms of a system of freely competitive markets is that the inherent difficulty in establishing certain markets for insurance brings about a sub-optimal allocation of resources. If we consider an investment, as does Hirshleifer, as an exchange of certain present income for uncertain future income, then the misallocation will take the form of under-investment.

Now consider Hirshleifer's recommendation that, in cases such as the one above,

a direct subsidy be used to induce more private investment rather than increase public investment. Suppose that a particular private investment were such that the benefits would be a marginal increase in the future supply of an existing commodity, i.e., this investment would neither introduce a new commodity nor affect future prices. Therefore, benefits can be measured at each point in time by the market value of this output, and can be fully captured through the sale of the commodity. Let $\bar{V}$ be the present value of expected net returns, and let V be the present value of net returns adjusted for risk where the certainty rate is used to discount both streams. Further, suppose there were a public investment, where the risks were publicly borne, for which the present value of expected net benefits was P. Since the risk is publicly borne, from the previous discussion it follows that P is the present value of net benefits adjusted for risk. Now suppose that $\bar{V}>P>V$. According to Hirshleifer, we should undertake the private investment rather than the public one, and pay a subsidy if necessary to induce private entrepreneurs to undertake this investment. Clearly, if there is a choice between one investment or the other, given the existing distribution of risk, the public investment is superior. The implication is that if a risky investment in the private sector is displaced by a public investment with a lower expected return but with a higher return when appropriate adjustments are made for risks, this represents a Hicks-Kaldor improvement. This is simply a restatement of the previous point that the government could more than pay the opportunity cost to the private entrepreneur.

Now consider the case for a direct subsidy to increase the level of private investment. One can only argue for direct subsidy of the private investment if $V<0<\bar{V}$. The minimum subsidy required is $|V|$. Suppose the taxpayers were to pay this subsidy, which is a transfer of income from the public at large to the private investor, in order to cover the loss from the investment. The net benefits, including the cost of risk-bearing, remain negative because while the subsidy has partially offset the cost of risk-bearing to the individual investor, it has not reduced this cost. Therefore, a direct public subsidy in this case results in a less efficient allocation of resources.

We can summarize as follows: It is implied by Hirshleifer that it is better to undertake an investment with a higher expected return than one with a lower expected return. (See 1965, p. 270.) This proposition is not in general valid, as the distribution of risk-bearing is critical. This statement is true, however, when the costs of risk-bearing associated with both investments are the same. What has been shown is that when risks are publicly borne, the costs of risk-bearing are negligible; therefore, a public investment with an expected return which is less than that of a given private investment may nevertheless be superior to the private alternative. Therefore, the fact that public investments with lower expected return may replace private investment is not necessarily cause for concern. Furthermore, a program of providing direct subsidies to encourage more private investment does not alter the costs of risk-bearing and, therefore, will encourage investments which are inefficient when the costs of risk are considered. The program which produces the desired result is one to insure private investments.

One might raise the question as to whether risk-spreading is not associated with large corporations so that the same result would apply, and it is easily seen that the same reasoning does apply. This can be made more precise by assuming there were n stockholders who were identical in the sense that their utility functions

were identical, their incomes were represented by identically distributed random variables, and they had the same share in the company. When the corporation undertakes an investment with a return in a given year represented by B, each stockholder's income is represented by $A+(1/n)B$. This assumes, of course, that a change in earnings was reflected in dividends, and that there were no business taxes. Clearly, this is identical to the situation previously described, and if n is large, the total cost of risk-bearing to the stockholders will be negligible. If the income or wealth of the stockholders were large with respect to the size of the investment, this result would be likely to hold. Note that whether or not the investment is a large one, with respect to the assets of the firm, is not relevant. While an investment may constitute a major part of a firm's assets if each stockholder's share in the firm is a small component of his income, the cost of risk-bearing to him will be very small. It then follows that if managers were acting in the interest of the firm's shareholders, they would essentially ignore risks and choose investments with the highest expected returns.

There are two important reasons why large corporations may behave as risk averters. First, in order to control the firm, some shareholder may hold a large block of stock which is a significant component of his wealth. If this were true, then, from his point of view, the costs of risk-bearing would not be negligible, and the firm should behave as a risk averter. Note in this case that the previous result does not hold because the cost of risk-bearing to each stockholder is not small, even though the number of stockholders is very large. Investment behavior in this case is essentially the same as the case of a single investor.

The second case is when, even though from the stockholder's point of view, risk should be ignored, it may not be in the interest of the corporate managers to neglect risk. Their careers and income are intimately related to the firm's performance. From their point of view, variations in the outcome of some corporate action impose very real costs. In this case, given a degree of autonomy, the corporate managers, in considering prospective investments, may discount for risk when it is not in the interest of the stockholders to do so.

Suppose that this were the case and also suppose that the marginal rate of time preference for each individual in the economy was 5 percent. From the point of view of the stockholders, risk can be ignored and any investment with an expected return which is greater than 5 percent should be undertaken. However, suppose that corporate managers discount for risk so that only investments with expected rates of return that exceed 10 percent are undertaken. From the point of view of the stockholders, the rate of return on these investments, taking risk into account, is over 10 percent. Given a marginal rate of time preference of 5 percent, it follows that from the point of view of the individual stockholder there is too little investment. Now suppose further that the government were considering an investment with an expected rate of return of 6 percent. Since the cost of risk-bearing is negligible, this investment should be undertaken since the marginal rate of time preference is less than 6 percent. However, in this case, if the financing were such that a private investment with a 10 percent expected rate of return is displaced by the public investment, there is a loss because in both cases the risk is distributed so as to make the total cost of risk-bearing negligible. The public investment should be undertaken, but only at the expense of consumption.

III. *The Actual Allocation of Risk*

In the idealized public investment considered in the last section, all benefits and costs accrued to the government and were distributed among the taxpayers. In this

sense, all uncertainty was borne collectively. Suppose instead that some benefits and costs of sizeable magnitudes accrued directly to individuals so that these individuals incurred the attendant costs of risk-bearing. In this case it is appropriate to discount for the risk, as would these individuals. Such a situation would arise in the case of a government irrigation project where the benefits accrued to farmers as increased income. The changes in farm income would be uncertain and, therefore, should be valued at more or less than their expected value, depending on the states in which they occur. If these increases were independent of other components of farm income, and if we assume that the farmer's utility were only a function of his income and not the state in which he receives that income, then he would value the investment project at less than the expected increase in his income, provided he is risk averse. If, however, the irrigation project paid out in periods of drought so that total farm income was not only increased but also stabilized, then the farmers would value the project at more than the expected increase in their incomes.

In general, some benefits and costs will accrue to the government and the uncertainties involved will be publicly borne; other benefits and costs will accrue to individuals and the attendant uncertainties will be borne privately. In the first case the cost of risk-bearing will be negligible; in the second case these costs may be significant. Therefore, in calculating the present value of returns from a public investment a distinction must be made between private and public benefits and costs. The present value of public benefits and costs should be evaluated by estimating the expected net benefits in each period and discounting them, using a discount factor appropriate for investments with certain returns. On the other hand, private benefits and costs must be discounted with respect to both time and risk in accordance with the preferences of the individuals to whom they accrue.

From the foregoing discussion it follows that different streams of benefits and costs should be treated in different ways with respect to uncertainty. One way to do this is to discount these streams of returns at different rates of discount ranging from the certainty rate for benefits and costs accruing to the government and using higher rates that reflect discounting for risk for returns accruing directly to individuals. Such a procedure raises some difficulties of identification, but this problem does not appear to be insurmountable. In general, costs are paid by the government, which receives some revenue, and the net stream should be discounted at a rate appropriate for certain returns. Benefits accruing directly to individuals should be discounted according to individual time and risk preferences. As a practical matter, Hirshleifer's suggestion of finding the marginal rate of return on assets with similar payoffs in the private sector, and using this as the rate of discount, appears reasonable for discounting those benefits and costs which accrue privately.

One problem arises with this latter procedure which has received little attention. In considering public investments, benefits and costs are aggregated and the discussion of uncertainty is carried out in terms of these aggregates. This obscures many of the uncertainties because benefits and costs do not in general accrue to the same individuals, and the attendant uncertainties should not be netted out when considering the totals. To make this clear, consider an investment where the benefits and costs varied greatly, depending on the state of nature, but where the difference between total benefits and total costs was constant for every state. Further, suppose that the benefits and costs accrued to different groups. While the investment is certain from a social point of view, there is considerable risk from a private point of view.

In the case of perfect markets for contingent claims, each individual will discount the stream of costs and benefits accruing to him at the appropriate rate for each time and state. However, suppose that such markets do not exist. Then risk-averse individuals will value the net benefits accruing to them at less than their expected value. Therefore, if net benefits accruing to this individual are positive, this requires discounting expected returns at a higher rate than that appropriate for certain returns. On the other hand, if net benefits to an individual are negative, this requires discounting expected returns at a rate lower than the certainty rate. Raising the rate of discount only reduces the present value of net benefits when they are positive. Therefore, the distinction must be made not only between benefits and costs which accrue to the public and those which accrue directly to individuals, but also between individuals whose net benefits are negative and those whose benefits are positive. If all benefits and costs accrued privately, and different individuals received the benefits and paid the costs, the appropriate procedure would be to discount the stream of expected benefits at a rate higher than the certainty rate, and costs at a rate lower than the certainty rate. This would hold even if the social totals were certain.

Fortunately, as a practical matter this may not be of great importance as most costs are borne publicly and, therefore, should be discounted using the certainty rate. Benefits often accrue to individuals, and where there are attendant uncertainties it is appropriate to discount the expected value of these benefits at higher rates, depending on the nature of the uncertainty and time-risk preferences of the individuals who receive these benefits. It is somewhat ironic that the practical implication of this analysis is that for the typical case where costs are borne publicly and benefits accrue privately, this procedure will qualify fewer projects than the procedure of using a higher rate to discount both benefits and costs.

REFERENCES

K. J. Arrow, "The Role of Securities in the Optimal Allocation of Risk-Bearing," *Rev. Econ. Stud.*, Apr. 1964, *31*, 91–96.

———, "Discounting and Public Investment Criteria," in A. V. Kneese and S. C. Smith, eds., *Water Research*. Baltimore 1966.

———, "The Economics of Moral Hazard: Further Comment," *Amer. Econ. Rev.*, June 1968, *58*, 537–38.

K. Borch, "The Safety Loading of Reinsurance," *Skandinavisk Aktuarietid-skrift*, 1960, 163–84.

G. Debreu, *Theory of Value*. New York 1959.

O. Eckstein, "A Survey of the Theory of Public Expenditure," and "Reply," *Public Finances: Needs, Sources, and Utilization*, Nat. Bur. Econ. Res., Princeton 1961, 493–504.

J. Hirshleifer, "Investment Decision under Uncertainty: Choice-Theoretic Approaches," *Quart. J. Econ.*, Nov. 1965, *79*, 509–36.

———, "Investment Decision under Uncertainty: Applications of the State-Preference Approach," *Quart. J. Econ.*, May 1966, *80*, 252–77.

———, J. C. De Haven, and J. W. Milliman, *Water Supply: Economics, Technology, and Policy*, Chicago 1960.

S. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quart. J. Econ.*, Feb. 1963, *77*, 95–111.

M. V. Pauly, "The Economics of Moral Hazard: Comment," *Amer. Econ. Rev.*, June 1968, *58*, 531–37.

P. A. Samuelson and W. Vickrey, "Discussion," *Amer. Econ. Rev. Proc.*, May 1964, *59*, 88–96.

J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 2d ed., New York 1964.

Hospital Cost Functions

By Judith R. Lave and Lester B. Lave*

The estimation of cost functions has received wide attention in recent years (see John Johnston, Alan Walters). Curve fitting techniques have given way to careful investigations of firm (or industry) models. Economists have begun to explore the possibilities that the cost function of a firm may shift over time (in a neutral or non-neutral fashion) and that a cross-section of firms may not all have the same cost function. More recently, investigators have attempted to account for errors of observation and the possibility of a simultaneous equations bias due to the level of cost affecting the rate of output (see Marc Nerlove).

All of these problems, with some added wrinkles, occur in estimating a cost function for a multi-product firm (see Ralph Pfouts). Even in the simplest terms, the multi-product firm requires a host of independent variables (one for each product) instead of a single one. Added to this difficulty is the likelihood that the quality of at least one product will change over time or that products will be differentiated across firms.

There are three problems that should be stressed. First, most econometricians would agree that an estimated cost function, rather than being a "true" specification of the transformed production function, is but an approximation over a relevant range. The nature of the approximation and the extent of the "relevant range" becomes more difficult to specify when there are many products. Hence, it seems even more advisable to regard the cost function of a multi-product firm as an approximation over a relevant range.

Second, regardless of whether one or many outputs is involved, estimated marginal costs tend to be sensitive to the specification of the function. This observation leads one to be cautious in estimating single product cost functions; it should lead one to be even more cautious in estimating multi-product cost functions. For such functions, it is essentially impossible to reflect the underlying cost structure in a single function (which is susceptible to estimation). The estimates from any multi-product cost function must be used with care.

The third difficulty arises when the measure of output is not a good one. For example, some firms (such as automobile manufacturers) produce a host of different advertised models and possible optional features. If each of the possible products were counted separately, it is inconceivable that a cost function could be estimated: There would be thousands of products. Either one must use a "composite output measure" or some measures of output mix.[1]

* The authors are instructor and associate professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University. This research was supported by a grant from the Public Health Service; the exploratory stages of the work were supported by Blue Cross of Western Pennsylvania. The opinions and any remaining errors are those of the authors. We wish to thank Martin Feldstein, Michael Lovell, Robert Lucas, Joseph Newhouse, Lester Taylor, and Alvin Cook for helpful suggestions, and Lester Silverman and Eugene Seskin for help with computations.

[1] The multi-product nature of output can be taken into account either by deriving a composite output measure which is a weighted sum of the various outputs or by including measures of the proportions of each output as explanatory variables. The former procedure simplifies estimation but depends on deriving a set of meaningful weights (for an example, see Franklin Fisher, et al.). In the latter procedure the estimated coefficient of each proportion is interpreted as being the relative marginal cost of producing that one type of

The hospital is a multi-product firm; it produces varying quantities of education, research, community services, outpatient care, and, its predominant activity, inpatient care. None of these activities is homogeneous across hospitals. In addition to the problems discussed above, the estimation of the hospital cost function is complicated by the difficulty of measuring the quality (appropriateness and competence) of each of the many services produced by the hospital.

The most commonly used measure of hospital output is unidimensional: the number of patient days produced.[2] The implicit assumption that all patient days are the same, both within and across hospitals, is obviously incorrect. In addition, the assumption is not a useful one in practice since the estimates of cost functions based on it have been characterized as nonsensical.[3] Several alternative measures have been suggested. However, much work remains to be done in defining and measuring the output of hospitals.

In this paper we develop a method for estimating the hospital cost function which attempts to deal with the problems stemming from the multi-product nature of output. With data consisting of time-series on many firms, the multi-product nature of each firm can be accounted for in either of two procedures; both of them are based on the assumption that, while the output mix differs among hospitals, it is constant within a hospital (over a short time period). The first procedure involves two stages of analysis and allows the cost function to differ among hospitals. In the first stage, a relation between average cost, utilization, size, and time is estimated for each hospital. The second stage is a search for the causes of the variation in the estimated parameters among hospitals. The second procedure is based on the more limited assumption that all hospitals have the same cost function. Here, the data are pooled (to form a cross-section time-series) and a single cost function is estimated.

Although we are concerned directly with hospitals, we believe that our techniques are applicable to other multi-product and multi-service industries. In Section I we describe the estimation procedure in detail. In Section II the results are presented, and in Section III the conclusions are discussed.

I. The Estimation Procedure

The Two-Stage Procedure

The cost per patient day (total cost divided by total patient days) may be written as in equation (1),

$$AC_{it} = f_{it}(U_{it}, S_{it}, X_{it}) \tag{1}$$

where U_{it} is utilization (output as a percent of capacity)[5] in period t, S_{it} is hospital size (the number of beds) in period t, and X_{it} is a vector which characterizes

product. Estimation is more difficult since the cost function must have a simple specification.

[2] A more relevant measure is the number of cases treated. M. Feldstein has used this measure in studying British hospitals. The question of whether patients or patient-days is the better measure cannot be settled a priori. The better measure is the one which is more homogeneous across hospitals, given the cost function specified. It was our judgment that the patient-day was more homogeneous, although there is little data with which to justify our belief.

[3] For a review of hospital cost literature, see J. Lave and J. Mann and D. Yett.

[4] One alternative to the patient-day involves counting the specific tasks performed by the hospital staff, e.g., lab tests, meals served, number of operations. But these tasks are inputs to the "restoration of health and alleviation of pain," not measures of the output itself, see M. Ingbar and L. Taylor. Another alternative, involving the construction of a composite measure of output has been undertaken by D. Saathoff and R. Kurtz and H. Cohen. The authors are currently investigating other approaches to composite output measures. Several attempts have been made to improve the patient-day as a measure. M. Feldstein, W. Carr and P. Feldstein and R. Berry have stratified by available facilities or corrected for case mix.

[5] This measure is often called the hospital occupancy rate and is measured as the ratio of total patient-days to available patient-days (bed days).

the product mix (both the diversity and quality of services provided) of hospital i in period t. As written, there is no presumption that the function remains constant over time in a given hospital or is the same across hospitals. So general a case is not considered. We do assume that the functional relations remain stable over short periods of time for each hospital but that neutral shifts in the function may occur.

As noted in the introduction, there is no satisfactory measure of hospital output. However, under special circumstances, the problem can be handled. We assume that the product mix of a hospital remains constant over a period of a few years.[6] If, as we assume, the hospital output vector is subject to proportional (or balanced) shifts, X remains unchanged over the period within a given hospital. If we were estimating a cost function for hospital i from time-series data, X_{it} could be dropped from equation (1), as shown in equation (2).

$$AC_{it} = A(t)f_i(U_{it}, S_{it}, e_{it}) \tag{2}$$

The $A(t)$ allows a neutral shift of the function over time. Insofar as output is not subject strictly to proportional expansion, an error e_{it} will be introduced. If there are additional variables (influencing the cost function) which are omitted, they would also contribute to an error term.

According to convention, short-run average (total) cost is a U-shaped function of output while the long-run average cost curve may display increasing, decreasing, or constant returns to scale. Most empirical studies of the short-run cost function have found an L-shaped relation (see Johnston and Walters). There are many reasons to expect an L-shaped curve for hospitals, the prime one being that hospitals tend to staff for a higher than average level of utilization so they can be ready on a stand-by basis. A functional form consistent with the L-shaped short-run average cost is the generalized Cobb-Douglas or linear in logarithms form. Another argument for this specification is that hospital costs have risen sharply in recent years. It seems likely that the variance of the error term in equation (2) is directly proportional to the size of AC. Thus, we specify the average cost curve for hospital i as in equation (3).[7]

$$\begin{aligned} \log AC_{it} = a_{i0} + a_{i1}t + a_{i2} \log U_{it} \\ + a_{i3} \log S_{it} + e_{it} \end{aligned} \tag{3}$$

With a number of observations on each hospital over time, the parameters of

[6] This hypothesized constancy of the product mix over a short period of time is being investigated by the authors. Some preliminary results suggest the hypothesis is a reasonable one, see Lave and Lave working paper. Any variation in case mix would probably be associated with a change in size; for our data the correlation between the log of hospital size and number of specialized facilities is .76. Thus, we can expect that the size variable would tend to pick up some of the variation in case mix.

[7] To determine the correct functional form, we tried a quadratic specification. No evidence could be found that the square of the utilization variable was relevant; the coefficient of size squared indicated that average cost declined slightly as size increased (there was no evidence that cost began to increase at the larger sizes). Thus, an L-shaped cost curve seemed to best describe the data.

There is no simultaneous equations problem in estimating hospital cost functions. The hospitals we study are nonprofit and accept all paying patients as long as there is space, i.e., hospitals do not choose their rate of output, but rather are constrained to accept all cases offered. In addition, it seems likely that the cost of the hospital has little effect on the demand for its services. Most patients have their expenses covered by insurance of some other "third party." In addition, one would presume the demand for hospital services is inelastic with respect to price.

Some evidence on these questions comes from looking at the correlation of the occupancy rate with cost. If patients chose among hospitals on the basis of price, one would expect to observe a negative correlation between occupancy rate and average cost; the correlation is positive and is .26. Secondly, one would expect that high cost hospitals would grow less rapidly than would low cost ones; the correlation between average cost and an index number of size (indicating the rate of growth of each hospital) is .37. Thus, what evidence there is supports the hypotheses.

equation (3) can be estimated for each hospital i. The coefficient of the time variable (t) indicates all changes in average cost which are geometrically related to time and are not accounted for by variation in utilization or size. Neutral technological change would contribute to a negative a_{i1}; increasing factor prices to a positive a_{i1}. In this model there is no way of separating cost reductions from cost increases and hence a_{i1} is a net inflation coefficient for each hospital. The coefficient of a_{i2} will indicate whether a hospital is operating on the declining portion of the short-run cost curve. Similarly, the coefficient of a_{i3} will indicate (if a hospital expanded) whether it captured economies of scale.

The output of this estimation procedure will be four estimated parameters for each hospital. The estimated relation for each hospital is a rather conventional time-series cost function. One might now ask: why do these parameters vary across hospitals? Is there any systematic way to explain differences in these parameters among hospitals? To answer these questions, a second-stage procedure is necessary.

Stage Two

Differences in hospital characteristics should lead to differences in the hospital cost function. Large hospitals tend to have more specialized facilities, the cost of which has been rising rapidly. They also tend to be located in urban areas where the price of labor inputs has been rising rapidly. However, large hospitals are generally run by trained managers who should be more efficient and better able to adapt to changing factor prices and technology. On balance, it is difficult to predict whether large hospitals should experience more or less inflation. Similarly, one can consider rural versus urban hospitals, or teaching versus nonteaching hospitals. Rural hospitals have lower absolute costs, but it is difficult to predict whether this gap is narrowing or widening over time. If, in fact, factor prices are rising more rapidly in the urban areas, one would predict that the gap is widening. Teaching programs tend to raise average costs. Hospitals with extensive teaching programs have the newest equipment. As medicine grows more expensive, these hospitals lead the way. We would expect that advanced teaching hospitals have had the highest rate of inflation, followed by teaching hospitals.

The management of a hospital tends to be permissive. Virtually all patients have their bills paid by insurance or government; even self-pay patients rarely worry about cost when hospitalization is required. The medical staff, which bears no responsibility for costs, generally makes all decisions about medical care. Possibly the only important check on cost is provided by a comparison of hospitals. If a hospital's costs rise more rapidly than those of comparable hospitals, some pressure will be exerted by Blue Cross, patients, and conscientious doctors. This behavioral observation provides the basis of a crude dynamic theory. Hospitals with relatively high initial cost will, *ceteris paribus*, tend to experience a slower rate of cost increase in the future.

This investigation of the variation of the a_{i1} can be formalized in a regression as shown in equation (4).

$$\begin{aligned} a_{i1} = {} & b_0 + b_1 \log S_{i0} + b_2 P_i + b_3 M_i \\ & + b_4 AT_i + b_5 T_i + b_6 \log AC_{i0} \\ & + b_7 \log U_{i0} + e_{it} \end{aligned} \tag{4}$$

Here S_{i0} is the size of the ith hospital in the first period, P_i and M_i are dummy variables indicating the hospital's location (Pittsburgh, other urban, or rural); AT_i and T_i are dummy variables representing

the hospital's teaching status (advanced teaching, teaching, or nonteaching), AC_{i0} and U_{i0} are, respectively, the hospital's average cost and utilization rate in the first period.

We also investigated the other parameters, a_{i2} and a_{i3}, using the explanatory variables of equation (4). We cannot, however, conjecture about the signs of the coefficients.

The One-Stage Procedure

In the two-stage procedure discussed above, we made the extreme assumption that the four parameters of equation (3) differ across hospitals. An assumption of the opposite extreme investigated here is that all four parameters are the same for each hospital. A simple time-series cross-section model expressing average cost as a function of time, utilization, and size is not plausible. There are, however, a number of compromises. One such model is expressed in equation (5).

$$\log AC_{it} = H_i + c_1 t + c_2 \log U_{it} + c_3 \log S_{it} + e_{it} \tag{5}$$

Here we assume that the coefficients of t, *log U*, and *log S* are identical across hospitals but that the intercepts differ. Allowing a different intercept H_i for each hospital means that some omitted variables can be accounted for. For example, if case-mix for each hospital is (as we assume) constant over short periods of time, it would add nothing to equation (5); the presence of H_i makes this relation more general than one with a measure of case-mix. This particular model implies that, whatever the level of absolute cost, a one percent change in U or S would give rise to the same percentage change in AC in all hospitals. Alternative models would allow any of the other three parameters, or combinations of the three parameters, to differ across hospitals.

It is straight forward to estimate models based on the extreme assumptions.[8] In the first instance, when all the parameters are assumed to vary across hospitals, a four variable equation is estimated for each hospital; in the second, when all parameters are assumed identical across hospitals, a single four variable relation is estimated. If one parameter is allowed to vary across hospitals, $n-1$ variables are added to the regression (where n is the number of hospitals in the analysis); if three parameters were allowed to differ, $3(n-1)$ independent variables would be added. Since there are 74 hospitals in our data, it is not trivial to allow the parameters to vary.

When this analysis was first undertaken, computational limitations prevented estimation of the more general models. Hence, the general model was modified by transforming AC, U, and S into index numbers as in equation (6),

$$\begin{aligned} I_{it} &= AC_{it}/\overline{AC}_i \\ J_{it} &= U_{it}/\overline{U}_i \\ K_{it} &= S_{it}/\overline{S}_i \end{aligned} \tag{6}$$

where $\overline{AC}_i$, $\overline{U}_i$, and $\overline{S}_i$ are the averages of the relevant variables for hospital i over the time-series. (Averages are taken rather than an initial observation in order to minimize dependence on a single observation.) The cost function is now specified in equation (7).

$$\log I_{it} = a_{i0} + a_{i1} t + a_{i2} \log J_{it} + a_{i3} \log K_{it} + e_{it} \tag{7}$$

That equation (7) is in fact similar to equation (3) is shown when the former is rewritten as equation (8).

[8] For the special case where the same explanatory variables are used for each hospital, A. Zellner points out that ordinary least squares applied to each equation in turn provides unbiased, minimum variance estimates. Since this is the case here, we have estimated the relation for each hospital separately.

$$\log AC_{it} = (a_{i0} + \log \overline{AC}_i - a_{i2} \log \overline{U}_i - a_{i3} \log \overline{S}_i) + a_{i1}t + a_{i2} \log U_{it} + a_{i3} \log S_{it} + e_{it} \tag{8}$$

If equations (3) and (7) were estimated using time-series data on a single hospital, the parameter estimates would be identical. If they were estimated using time-series data pooled from a number of hospitals, the parameter estimates would be quite different. Equation (7) allows each hospital to have its own intercept. The intercepts are not completely free to vary as would be true if separate dummy variables were used (as in equation (5)), but they do allow variation based on each hospital's average values for cost, utilization, and size.

In formalizing a cross-section time-series model, it seemed reasonable to assume that certain hospital characteristics may be used as surrogates for other factors affecting hospital cost. As noted above in the discussion of the second stage of the two-stage procedure, we would expect that part of the change in hospital cost could be accounted for by the hospital's teaching status, its location, and initial values of size, utilization, and average cost. Hence, one function to be estimated from the pooled cross-section time-series data is shown in equation (9).

$$\log I_{it} = d_0 + d_1 t + d_2 \log J_{it} + d_3 \log K_{it} + d_4\, AT_i + d_5 T_i + d_6 P_i + d_7 M_i + d_8 \log AC_{i0} + d_9 \log U_{i0} + d_{10} \log S_{i0} + e \tag{9}$$

We hypothesized that d_2 and d_3 are negative, indicating decreasing short- and long-run costs; d_4, d_5, d_6, and d_7 are positive indicating that urban, advanced teaching hospitals were more expensive than their rural, nonteaching competitors; d_8 is negative indicating that hospitals with high initial cost had less rapid cost increases; d_{10} is positive indicating that large hospitals had greater cost increases; we did not conjecture on d_9.

The Data: Some Difficulties

Data were obtained on 74 western Pennsylvania hospitals for the period 1961–67. In all, 14 semiannual observations were obtained on each hospital. Cost and utilization data were obtained from *Comparative Financial and Statistical Information*, a report by Blue Cross of Western Pennsylvania. Pennsylvania hospitals are required to submit semiannual reports to the Insurance Commissioner of Pennsylvania. These reports are checked and summarized by the audit staff of Blue Cross. Data on hospital size, number of facilities, geographical location, and teaching status were obtained from unpublished data series by Katherine Barker of Blue Cross of Western Pennsylvania. Her data on hospital size and number of facilities were based on information published in the annual August issue of *Hospitals, Journal of the American Hospital Association.*

After the analysis described in Section I was performed, we went on to extend the results and corroborate them with data from Blue Cross of Greater Philadelphia. When the Pittsburgh data were aggregated into annual observations, a consistent result was that all coefficients agreed with the semiannual analysis except that of utilization; it fell from −.6 to −.3. The latter value was nearly identical to the parameter estimates from the Philadelphia (annual) data. In all cases, the estimated parameter was extremely significant (more than ten times its standard error). The consistency between Pittsburgh and Philadelphia results implied that −.3 was the correct estimate. However the inconsistency between semiannual and annual results (for Pittsburgh data) was disconcerting since it implied that the underlying

structure was not compatible with our model.

An inquiry into data collection revealed that the first semiannual observation (for each hospital) is an estimate, unaudited by Blue Cross. At the end of the year, Blue Cross subjects the annual estimate to an intense audit to determine costs for purposes of reimbursement. The second semiannual observation is then calculated as the difference between the audited annual figure and the estimated first half costs, i.e., it is a residual. Questions were raised about the accuracy of the first semiannual cost estimates.

We investigated the possibility of errors in the semiannual figures by inserting a dummy variable for first or second half of the year (F takes on a value of 0 if the observation is for the first half and a value of 1 if it is for the second half). The correlation between F and utilization is $-.5$ indicating that the occupancy rate is significantly higher in the first half than it is in the second. When F is inserted in the regression, all parameter estimates are nearly identical to those from the annual data.

The apparent explanation is that, since costs have been rising rapidly (more than 3 percent annually for these hospitals), hospitals tend to underestimate first half costs systematically (by about 2 percent and thus overestimate second half costs by 2 percent). When this underestimate is combined with the systematic difference in utilization, the estimate of marginal cost is biased downward. The artificially low average cost of the first half occurs when there is high utilization; the artificially high average cost of the second half occurs when there is low utilization. Thus, the cost increase due to higher utilization is systematically underestimated. The only parameter affected by adding F to the regressions is the coefficient of utilization; we present the results of the extended model in the sections which follow.

II. *The Results*

The Two-Stage Procedure

In the first stage of the two-stage procedure, we estimated equation (3), with F added for each of the 14 hospitals in our sample. With 14 observations, there were 9 degrees of freedom. The results were in accordance with our expectations and the coefficients of determination were almost always greater than .9.

The coefficients of the time dummy variable, a_{i1}, had a mean of 3.28 percent (indicating a mean rate of inflation of 3.28 percent) and a standard deviation of 1.04 percent. This coefficient was almost invariably significant and it ranged from .7 percent to 6.0 percent. A histogram showed the distribution of the coefficient to be unimodal with its median almost identical to the mean.

The coefficient of utilization a_{i2} had a mean of $-.605$ and a standard deviation of .591. Thus, on average, variable cost was 40 percent of total cost. The coefficient was generally significant with a range from -2.775 to .893. A histogram showed the distribution of this coefficient to be unimodal with its median approximately equal to its mean.

The coefficient of size, a_{i3}, had a mean of $-.333$ and a standard deviation of 2.418. The coefficient was almost never significant with a range from -13.4 to 8.5. Although the distribution of the coefficient was unimodal, the median was about $-.03$. Very little can be said about the slope of the long-run average cost curve from these results. The coefficient of period had both mean and standard deviations equal to .020; it ranged from $-.029$ to .100 with a single mode and median at .020.

Stage two, where we were attempting to explain the variation in a coefficient among hospitals, was frankly a search for significant coefficients. We were attempting to

Table 1—Hospital Cost Increases[a,b]

a_{i1} =	b_0	$+b_1 log AC_0$	$+b_2 log S_0$	$+ b_3 log U_0$	$+ b_4 AT$	$+ b_5 T$	$+ b_6 P$	$+ b_7 M$	$+ b_8 log K_0$	$\bar{R}_2$
(1−1)	.240	−.047	.024	0.035						.210
		(2.41)	(4.67)	(−1.58)						
(1−2)	.272	−.052		−.024	.013	.009				.154
		(−2.46)		(−1.06)	(3.68)	(3.41)				
(1−3)	.178	−.036		−.007			.004	−.005		.047
		(1.52)		(−.34)			(1.15)	(1.61)		
(1−4)	.294	−.058	.016	−.036	.004	.005	.001	−.003		.213
		(−2.59)	(2.14)	(1.60)	(.78)	(1.40)	(.20)	(−.25)		
(1−5)	.659	−.060	.021	−.058	.002	.004	.002	−.003	−.061	.235
		(−2.72)	(2.60)	(2.27)	(.43)	(1.24)	(.54)	(−.94)	(−1.70)	

[a] Values of the t-statistic are shown in parentheses under the coefficients, in Tables 1–4.
[b] There were 74 observations:

Variable	Mean	Variable	Mean	Variable	Mean
a_{i1}	.033	$log\ U_0$	2.888	P	.324
$log\ AC_0$	3.406	AT	.208	M	.338
$log\ S_0$	2.249	T	.364	$log\ K_0$	4.975

learn whether the variation in each of the coefficients could be explained by hospital characteristics, how much of the variance could be explained, and which variables had an affect. Many explanatory variables were tried.

Hospital Cost Increases

Five regressions, reported in Table 1, were used to explain the variance in the inflation coefficient across hospitals. The dependent variable, a_{i1}, is the rate of cost increase for hospital i, after accounting for cost changes due to variation in utilization or size. These regressions explain nearly one-fourth of the variation in a_{i1}, correcting for the loss in degrees of freedom. In the first regression the inflation coefficient for each hospital is a function of its initial cost per patient day (AC_0), its initial size (S_0) and its initial utilization rate (U_0). The signs of the coefficients indicate that, *ceteris paribus*, hospitals with high initial average costs tended to have relatively smaller rates of cost increase; large hospitals tended to have relatively larger rates of cost increase; and hospitals with high initial utilization tended to have relatively smaller rates of cost increase. Both the coefficients of AC_0 and S_0 are significant; the coefficient of U_0 approaches significance.

The signs of the regression coefficients accord with our expectations. Large hospitals tend to have teaching programs, more specialized treatment facilities, and to be located in large cities. We speculate that the costs of teaching programs and specialized care have risen more rapidly than the cost of simple care. We also speculate that urban wage rates have risen relative to rural wage rates. All three of these factors would tend to account for the positive coefficient of S_0. These data do not permit investigation of our speculations; we intend to pursue them in future research.

The negative coefficient for U_0 is difficult to interpret. Since it is widely recognized that short-run marginal cost is a fraction of average cost, hospital administrators realize that one way of holding down cost is to keep the utilization rate high. Thus, the negative coefficient for U_0 might be taken

to mean that hospitals with high utilization are, *ceteris paribus*, hospitals with a concern for holding down costs.

Regressions (1–2), (1–3), (1–4), and (1–5) contain variables indicating the nature of a hospital's teaching program and location. A hospital is classified as having an advanced teaching program (AT), a teaching program (T), or no teaching program. Three dummy variables can be used to describe these classifications. Note that one variable must be excluded to fit the regression. The excluded variable (those hospitals with no teaching program) has a coefficient of zero, relative to hospitals with teaching programs (actually, the coefficient of the excluded variable is part of the constant term). The location of a hospital is also defined in terms of three dummy variables. If a hospital is located in Pittsburgh or the surrounding counties, it is classified as P. If it is in one of the other western Pennsylvania counties which has at least 50 percent of its residents in cities, it is classified as M. Otherwise, it is classified as rural and, relative to P and M hospitals, has a coefficient of zero.

In regression (1–2) the coefficients of AT and T indicate that hospitals with advanced teaching programs averaged a 1.3 percentage point (per year) higher rate of cost inflation than hospitals with no teaching programs; hospitals with regular teaching programs averaged a 0.9 percentage point (per year) higher inflation than hospitals with no teaching programs. In regression (1–4), when variables representing hospital location are included, the size of the teaching coefficients decreases (as would be expected). The coefficients, however, still indicate that the rate of cost inflation was higher in advanced teaching hospitals than in teaching hospitals, and higher in teaching hospitals than in nonteaching hospitals.

In regression (1–3) the coefficient of P and M indicate that hospitals in Pittsburgh had a .4 percentage point greater increase in cost than rural hospitals; Metropolitan hospitals had a .5 percentage point lower rate of cost increase than rural hospitals. Since neither M nor P is significant, there is no conclusive evidence for our conjecture that rural hospitals probably experienced a lower rate of cost inflation than urban hospitals.

In the final regression, the initial value of the size index number is added. If hospitals grew rapidly, their value of K_0 will be quite small; hospitals that did not grow will have a K_0 of unity. The negative coefficient for K_0 indicates that hospitals which grew most rapidly had the greatest rate of cost increase; hospitals which did not grow at all had the smallest rate of cost increase. However, this effect is not significant.

The Utilization Coefficient

The regressions analyzing the variation in a_{i2} the coefficients of the utilization index are reported in Table 2. (The coefficient of the utilization index carries the interpretation that a 1 percentage point increase in mean utilization implies an a_{i2} percentage point decrease in mean average cost.) In interpreting the results of the regressions it should be remembered that, since the utilization coefficient is negative, a small coefficient is one which is large in absolute value.

In the first regression the independent variable is K_0, the initial value of the size index. The negative sign indicates that hospitals which experienced the greatest increase in size had the largest a_{i2} (smallest a_{i2} in absolute value). That is, hospitals experiencing the greatest increase in size had the greatest increase in average cost when utilization increased. In other words, the hospitals which grew the most had the highest marginal cost.[9]

[9] To be precise, hospitals which grew the most had total costs rise most (as a percentage increase) when utilization increased. One must be careful in making the statement above in the text since a small sign might

TABLE 2—VARIATION IN UTILIZATION COEFFICIENT

a_{i2} =	b_0	+ $b_1 \log K_0$	+ $b_2 \log S_0$	+ $b_3 \log U_0$	+ $b_4 \log AC_0$	+ $b_5 AT$	+ $b_6 T$	+ $b_7 P$	+ $b_8 M$	$\bar{R}_2$
(2–1)	16.669	−3.472 (−1.80)								.030
(2–2)	15.162	−3.128 (−1.67)						−.184 (−1.14)	−.429 (−2.69)	.096
(2–3)	14.784	−3.041 (−1.62)				.037 (.18)	−.183 (−1.16)	−.173 (−.93)	−.430 (−2.70)	.098
(2–4)	−6.329			2.846 (2.18)	−.640 (−.48)	−.002 (−.01)	−.237 (−1.47)	−.225 (−1.12)	−.461 (−2.91)	.115
(2–5)	3.219	−1.596 (− .72)	.009 (.02)	2.335 (1.48)	−.691 (−.51)	.005 (.02)	−.225 (−1.12)	−.195 (−.94)	−.446 (−2.70)	.096

There were 74 observations: mean $a_{i2} = -.605$, mean $a_{i3} = -.333$

The second regression adds variables for the location of each hospital. The negative coefficients mean that urban hospitals have lower marginal cost than rural ones. Teaching status is added in regression (2–3). There is some indication that teaching hospitals have lower marginal cost than either advanced teaching or nonteaching hospitals.

The fourth regression substitutes initial utilization and average cost for K_0. The significance of location and teaching variables is enhanced and the coefficient of determination is increased. The positive, significant coefficient of initial utilization is interpreted to mean that hospitals with high utilization have high marginal cost, i.e., increases in utilization do not lower average cost much.

Regression (2–5) collects all of the variables in a single regression. Initial size is of no significance and K_0 loses its significance in the presence of other variables. The adjusted coefficient of determination falls when these additional variables are entered.

indicate something other than a large marginal cost. For example, if two hospitals had the same (dollar value of) marginal cost, the one with lower average cost would have its total cost rise more rapidly as utilization increased.

Variations in First-Half Coefficient

Our attempts to explain the variation in the first-half coefficient a_{i4} are presented in Table 3. This coefficient is an estimate of the extent to which first-half average cost was systematically underestimated by a hospital; the mean underestimate is 2 percent. Regression (3–1) indicates that large hospitals were more guilty of underestimation than small ones, that hospitals with high initial utilization and low initial average cost were also most guilty of underestimation. Regression (3–2) adds location variables. Not surprisingly, urban hospitals are less guilty of cost underestimation than rural hospitals. Teaching variables are substituted for location variables in regression (3–3) and the coefficients indicate that advanced teaching hospitals are more guilty of underestimation than are other hospitals. Regression (3–4) combines the variables of the last two regressions and has an adjusted coefficient of determination greater than .22; the coefficients are unchanged in the combined regression. In the final regression, K_0 and S_0 are added with no effect; the other coefficients are unchanged and the adjusted coefficient of determination falls due to the loss in degrees of freedom.

TABLE 3—VARIATION IN FIRST-HALF COEFFICIENT

a_{i4} =	a_0	$+a_1 logK_0$	$+a_2 logS_0$	$+a_3 logU_0$	$+a_4 logAC_0$	$+a_5 AT$	$+a_6 T$	$+a_7 P$	$+a_8 M$	$\bar{R}^2$
(3−1)	.269		.012 (1.15)	.041 (.87)	−.116 (−2.77)					.070
(3−2)	.228			.058 (1.35)	−.108 (−2.49)			−.004 (−.59)	−.017 (−3.14)	.169
(3−3)	.270			.055 (1.24)	−.120 (−2.84)	.012 (1.68)	−.005 (.94)			.128
(3−4)	.244			.060 (1.43)	−.115 (−2.64)	.007 (1.37)	−.006 (−1.19)	−.006 (−.94)	−.017 (−3.24)	.226
(3−5)	.277	−.006 (−.08)	−.002 (−.12)	.060 (1.17)	−.115 (−2.60)	.010 (1.04)	−.006 (−.88)	−.006 (−.89)	−.017 (−3.10)	.203

There were 74 observations: mean a_{i4} = .020

Our attempts to analyze the a_{i3} proved completely fruitless. No variable had a significant simple or partial correlation coefficient. Since few of the a_{i3} were significant, we can only conclude that the size coefficients appear to be random numbers.

The One-Stage Procedure

In the one-stage procedure, we pooled the data of all hospitals and estimated equation (7). The estimated equation is presented as regression (4–1) in Table 4. Before continuing the analysis of the one-stage procedure, we determined whether or not the data should be pooled at all.

In estimating the 74 equations, one determines best linear, unbiased estimates of the parameters for each hospital. There is, however, little generality in this host of cost functions. In pooling the data, one determines the best estimated parameters for the combination of hospitals. However, one must be careful in pooling data. If the individual cost functions are not identical, the pooled regression can be useless in that it fails to explain a significant amount of variation or it can be misleading in that its coefficients are quite different from the coefficients estimated in the individual regressions.

The $\bar{R}^2$ of .855 for regression (4–1) indicates that the pooled data fit the relation well. The hypothesis that the pooled regression explains as much variation after accounting for the degrees of freedom can be tested using the analysis of covariance.[10] The F value of 4.05 with 365 and 666 degrees of freedom is significant at conventional levels. Thus, we must formally conclude that we do lose explanatory power by forcing all five estimated parameters to be identical across hospitals.

We then investigated the possibility that only one of the parameters differed across hospitals (the other three being held constant). Five regressions were estimated in this manner. These were compared to the regression wherein all five parameters were assumed to be identical. Allowing a parameter to vary did not contribute to the explanatory power of the regression. In each case the F value was less than one. It did not seem profitable (and was computationally impossible) to investigate the hypotheses that two or more parameters differ across hospitals while the remaining parameters were constant.

We also compared the parameters esti-

[10] The presence of autocorrelation in the data would tend to bias the F test toward indicating a significant difference where none existed. However, this statistic is a robust one and unlikely to be affected much.

TABLE 4—COST REGRESSION WITH POOLED DATA

(4–1)
$$\log I = 6.670 \underset{(-11.41)}{-.357 \log J} \underset{(-.22)}{-.006 \log K} \underset{(62.19)}{+.031t} \underset{(14.81)}{+.028F} \qquad \bar{R}^2 = .855$$

(4–2)
$$= 6.503 \underset{(-11.74)}{-.317 \log J} \underset{(-.15)}{-.003 \log K} \underset{(8.05)}{+.021t_1} \underset{(16.18)}{+.042t_2} \underset{(24.98)}{+.067t_3} \underset{(34.03)}{+.096t_4} \underset{(45.36)}{+.130t_5} \underset{(66.56)}{+.195t_6} \underset{(18.03)}{+.029F} \qquad \bar{R}^2 = .894$$

(4–3)
$$= 6.638 \underset{(-12.42)}{-.334 \log J} \underset{(-.33)}{-.007 \log K} \underset{(3.18)}{+.045t_1} \underset{(3.25)}{+.091t_2} \underset{(3.33)}{+.139t_3} \underset{(3.45)}{+.192t_4} \underset{(3.60)}{+.250t_5} \underset{(4.07)}{+.340t_6} \underset{(-2.63)}{-.008t(\log AC_0)}$$

$$\underset{(.33)}{+.001t(\log U_0)} \underset{(.32)}{+.002t(\log S_0)} \underset{(7.45)}{+.007t(AT)} \underset{(3.14)}{+.002t(T)} \underset{(-5.95)}{-.041AT(\log S_0)} \underset{(-3.60)}{-.006T(\log S_0)} \underset{(1.56)}{+.008AT(P)}$$

$$\underset{(2.08)}{+.013AT(M)} \underset{(2.27)}{+.009T(P)} \underset{(2.43)}{+.008T(M)} \underset{(-1.70)}{-.001t(P)} \underset{(-3.64)}{-.002t(M)} \underset{(2.24)}{+.025F} \qquad \bar{R}^2 = .900$$

(4–4)
$$= 5.928 \underset{(-11.67)}{-.323 \log J} \underset{(.10)}{+.002 \log K} \underset{(6.73)}{+.199t_1} \underset{(6.76)}{+.398t_2} \underset{(6.80)}{+.600t_3} \underset{(6.87)}{+.807t_4} \underset{(6.94)}{+1.019t_5} \underset{(7.16)}{+1.262t_6} \underset{(8.34)}{+.228 \log AC}$$

$$\underset{(-.14)}{-.005 \log U_0} \underset{(.53)}{+.003 \log S_0} \underset{(10.03)}{+.010t(AT)} \underset{(5.16)}{+.004t(T)} \underset{(-.58)}{-.001t(P)} \underset{(-2.37)}{-.001t(M)} \underset{(-8.52)}{-.054t(\log AC_0)}$$

$$\underset{(.05)}{+.001t(\log U_0)} \underset{(.05)}{+.001t(\log S_0)} \underset{(-8.42)}{-.018AT(\log S_0)} \underset{(-4.71)}{-.008T(\log S_0)} \underset{(2.33)}{+.028F} \qquad \bar{R}^2 = .906$$

mated in the individual and pooled regressions. The coefficient of t in the pooled regression is 3.1 percent, the mean of the 74 time coefficients is 3.3 percent. These two values are quite comparable. The coefficient of J in the pooled regression is $-.357$, the mean of the 74 coefficients is $-.605$. These two values are only roughly comparable. The single coefficient indicates that 64 percent of average cost is variable with utilization while the mean of the coefficients suggest that 40 percent of average cost is variable. If we were to choose between these estimates, we would choose the former. We should point out that in the individual regressions the coefficient of $\log U_{it}$ was less than -1.0 in 16 instances, indicating that total cost fell as more patients were treated, and positive in 12 instances; this raises problems as to the plausibility of the results. The coefficient of K in the pooled regression is $-.006$, the mean of the 74 coefficients is $-.33$. Since size is so unimportant in these regressions, rarely attaining significance, these coefficients are not too dissimilar. The coefficient of first half is .028 in the pooled regression while the mean of the individual regressions is quite close at .020. We concluded that pooling the data does not represent a great loss in explanatory power and does add to the possibility of richer models since the number of observations increases.

We argued in Section I that additional hospital characteristics (other than size and utilization) could be added to the cost function. Before proceeding, we investigated whether it would be fruitful to disaggregate the time variable. A regression in which six annual dummy variables substitute for the linear time variable is reported as regression (4–2). The intercept of this estimated cost function is 6.503 in 1961, 6.524 in 1962, 6.545 in 1963, etc. The adjusted coefficient of determination is much higher for this regression than for regression (4–1) and an F test shows that the explanatory power of this regression is significantly increased. The indicated average rate of inflation is 2.8 percent per year and the coefficients of the dummy variables indicate that the rate of cost increase has been rising over time: the first two years average only 2.1 percent; the third year, 2.5 percent; the fourth, 2.9 percent; the fifth, 3.4 percent; and the sixth year, 6.5 percent.

We then added a number of independent variables, representing various hospital characteristics, to the above regression. Taking a clue from the regressions reported in Tables 1 and 2, we also added interaction variables. Each of the independent variables was multiplied by the linear time variable to produce an additional set of explanatory variables. (Multiplying a variable by t means that it affects the increase in cost over time rather than the level of cost.) It also seemed possible that teaching hospitals might experience different rates of cost increase depending on their location and size. We defined six more interaction variables by taking the products of the dummy variables for teaching and location, and the product of the teaching dummy vriables and the log of initial size. These variables are included in regressions (4–3) and (4–4).

The coefficients of J_{it} and K_{it} in regression (4–3) are almost totally unaffected by the addition of new variables. The coefficients of the time dummy variables increase significantly. Apparently, the average hospital experienced a rate of cost increase of approximately 4.9 percent, when the special rates of cost increase associated with size, teaching status, location, initial cost and initial utilization are held constant. There do seem to be significant interactions among the variables. The interpretation of the negative coefficients for $t(\log AC_{i0})$ is that hospitals with a high initial cost had their costs increase less rapidly over time; in each year their costs

fell relative to hospitals with a low initial average cost. Thus, a high initial average cost resulted in a long term adjustment to bring costs into line with other hospitals. The coefficients for t($logU_{i0}$) and t($logS_{i0}$) are completely insignificant. The coefficients for t(AT) and t(T) mean that advanced teaching hospitals experienced a higher rate of cost increase than teaching hospitals which, in turn, had a higher cost increase than nonteaching hospitals.

There is mixed evidence on the relative rate of cost increase in large hospitals. The coefficient of t($logS_{i0}$) is completely insignificant. The coefficients of $AT(logS_{i0})$ and $T(logS_{i0})$ indicate that large teaching hospitals experienced a lower rate of cost increase than did small teaching hospitals; furthermore, economies of scale are suggested to be more important for advanced teaching hospitals than for teaching hospitals. The negative coefficients of t(P) and t(M) indicate that rural hospitals experienced higher cost increases than urban hospitals.

In regression (4–4) we added variables for initial average cost, utilization, and size, dropping variables for the interaction of teaching and location. These changes caused the coefficient of K_{it} to become positive (and completely insignificant) and the coefficients of time to rise (indicating an increased rate of inflation). The coefficients of one of the added variables is significant, implying that the functional form of the previous regression can be improved. Thus, hospitals with high initial cost had a much larger than normal cost increase in 1961, but a much smaller than normal cost increase in 1967. Previously we concluded that hospitals with initial high cost had slower rates of cost increase in all years; now we conclude that they had higher rates of cost increase during 1961–64 and lower rates of cost increase during 1965–67. Initial utilization and size add nothing to the regression.

To examine the effect of collinearity, a final regression was run which included all variables.[11] The adjusted coefficient of determination was .910 and the coefficients were similar to regression (4–4). The coefficient of K_{it} remained positive and the rate of inflation was extremely high ($t_6=2.445$).

Comparison of the Estimation Procedures

While it is difficult to compare the results of the one-stage procedure with those of the two-stage procedure, it is possible to check their consistency. The mean values of the estimated coefficients for the time, size, and utilization variables in the 74 individual hospital regressions have already been compared with the corresponding coefficients for the pooled regressions. It was concluded that they were similar. We now examine the results of the second stage regression to compare them with those obtained by inserting the interaction variables into the pooled regression.

In analyzing the regressions in Table 1 (the analysis of the variation in the net inflation coefficient a_{i1}) we determined that hospitals with high initial cost and those with high initial utilization had lower rates of inflation. Large teaching hospitals had a greater rate of cost increase than nonteaching hospitals (with advanced teaching hospitals leading). There was mixed evidence on the affect of location.

These conclusions are generally borne out in regression (4–4). Hospitals with high initial average cost experienced significantly slower rates of cost increase. Advanced teaching hospitals experienced the greatest rates of cost increase followed by teaching hospitals. Urban hospitals experienced lower rates of cost increase than did rural hospitals, thus contradicting our hypothesis and settling the inconclusive nature of regression (1–5). There are no contradictions to the other results in re-

[11] The results of this regression are not shown, but will be made available by the authors, upon request.

gression (4–5), but the other results are not corroborated because the coefficients are completely insignificant. We conclude that the two estimation procedures have the same implications.

As explained in Section I, transforming the data into index numbers allowed each hospital to have its own intercept, based on its average size, utilization, and average cost. The generalization of this transformation is to allow each hospital to have a separate intercept. Our hypothesis is that if the variables are transformed into index numbers, there is no additional gain in allowing each hospital to have its own intercept. A statistical test of this hypothesis is an F test on whether adding 73 dummy variables increases the explanatory power of the regression. Since the value of F is less than 1.0, we conclude that nothing is added by allowing a different intercept for each hospital. A more sensitive test of the hypothesis is to examine the estimated parameters of the two regressions. We find that the coefficients of time, utilization, and size are nearly identical. This result suggests that the specification is valid.

III. *Summary and Conclusion*

In applying our method for estimating cost functions of multi-product firms to the hospital industry, much of our investigation consisted of a search for significant coefficients. This search was made necessary because there is little knowledge of the functional form of the cost relationship or of which explanatory variables to include. We were able to specify a number of signs, but there were many that eluded prediction. For these reasons, our conclusions should be viewed as tentative. It should be noted that at the time this analysis was first undertaken we had available only nine observations on each hospital. With one exception the results of the preliminary analysis were remarkably similar to results reported here using the extended data. To our surprise, the variables used in the pooled regressions continued to have similar signs and magnitudes; occasionally some of the significance tests changed.[12] Another significant change in specification came when the first-half dummy variable was added. With the exception of the utilization coefficient, adding this variable had almost no affect on the results. Further evidence is provided by comparing these results to those derived from replicating the procedure with annual observations for Pittsburgh and Philadelphia hospitals. Even the interaction variables continue to display similar coefficients.[13]

Our method of accounting for the multiproduct nature of the hospital is computationally simple and plausible. Some elaborate methods have been tried and it is of some interest to compare results.

One question that has received much attention is the relationship between marginal and average cost. The commonly held belief is that an empty bed is about 75 percent as expensive as a full one, i.e., mar-

[12] The only important change in the results reported in Table 1 occurs in the coefficient of M. In the first estimation, M had a positive sign and was about one-third the magnitude of the coefficient of P; these results were much more in line with expectations than the negative M shown in Table 1. Our previous estimation converted a clerical error into a positive coefficient for $\log K^0$ in Table 2. As in our current Table 2, the coefficient of AT continued to be implausible in some regressions. In the former estimation, we were able to explain about 10 percent of the variation in a_{i3}; the range of estimated coefficients was considerably smaller, from -3.50 to 4.46. Apparently the shorter time-series generated spurious significance. In the pooled regression, any time a coefficient was significant, neither the sign nor magnitude changed as more observations were added. Previously, the coefficient of $\log K_{i0}$ tended to be positive and insignificant; now it is negative and insignificant. In the first estimation, the interaction variables were not so stable in sign; t($\log S_{i0}$) generally had a negative (insignificant) coefficient; t($\log S_{i0}$) sometimes had a negative coefficient.

[13] Space limitations prevent us from presenting both Pittsburgh and Philadelphia annual results here; we do this in our forthcoming paper in *Inquiry*.

ginal cost is a small percentage of average cost. Our results suggest that the short-run average cost curve is L-shaped and that marginal cost is a large percentage of average cost; we estimate that marginal cost is between 40 and 65 percent of average cost. These results disagree with those of P. Feldstein who found that only 20 percent of cost is variable. There is no real contradiction here since his results are for one month periods while ours are for six month periods; one would expect that a higher percentage of cost is variable as the length of the period grows long. Our results are supported by the analyses of M. Feldstein and Ingbar and Taylor. In analyzing cost per patient week, Feldstein finds that marginal cost is approximately equal to average cost (although the focus of his book is an analysis of cost per patient). Ingbar and Taylor estimate a similar relation (not transformed into logarithms)and find that average cost is affected little by changes in utilization. Although it is a bit difficult to come up with a single figure for percentage variable, their results suggest that marginal cost is about 80 percent of average cost.

Second, our results indicate that if economies of scale exist in the hospital industry, they are not very strong. In most cases the sign of our K variable was negative which means that as a hospital grew it captured some economies of scale. The sign, however, rarely attained significance, which accords with the results of Ingbar and Taylor, Berry, and M. Feldstein.

Thirdly, our results suggest that the rate of cost increase has been accelerating (as would be indicated by conventional price indices): the cost per patient day increased 2.1 percent in 1961 and 6.5 percent in 1967 (the first year of Medicare). We would add that conventional price indices should *understate* the rate of cost increase since both size and utilization increased in western Pennsylvania hospitals over the period and their coefficients were negative.

We conclude that the proposed method of estimating hospital costs seems fruitful. However, one important problem not answered is that of determining the absolute level of cost for a hospital; we have concentrated on explaining changes in cost. The determination of the level of cost is an important problem, and the subject of future research.

REFERENCES

R. Berry, "Returns to Scale in the Production of Hospital Services," *Health Service Research,* Summer 1967, *2,* 123–39.

W. Carr and P. Feldstein, "The Relationship of Cost to Hospital Size," *Inquiry,* Mar. 1967, *4,* 45–65.

H. Cohen, "Variations in Cost Among Hospitals of Different Sizes," *Southern Econ. J.,* Jan. 1967, *33,* 355–66.

M. Feldstein, *Economic Analysis for Health Service Efficiency,* Amsterdam 1967.

P. Feldstein, *An Empirical Investigation of the Marginal Cost of Hospital Services,* Chicago 1961.

F. Fisher, Z. Griliches, and C. Kaysen, "The Costs of Automobile Model Changes since 1949," *J. Polit. Econ.,* Oct. 1962, *70,* 433–51.

M. Ingbar and L. Taylor, *Hospital Costs in Massachusetts,* Cambridge, Mass. 1968.

J. Johnston, *Statistical Cost Analysis,* New York 1960.

J. Lave, "A Review of the Methods Used to Study Hospital Costs," *Inquiry,* May 1966, *3,* 57–81.

——— and L. Lave, "Estimated Cost Relations for Pennsylvania Hospitals," *Inquiry,* 1970 forthcoming.

——— and ———, "The Extent of Role Differentiation Among Hospitals," working paper, Carnegie-Mellon Univ., Apr. 1970.

J. Mann and D. Yett, "The Analysis of Hospital Costs: A Review Article," *J. Bus., Univ. Chicago,* Apr. 1968, *41,* 191–202.

M. Nerlove, *Estimation and Identification of*

Cobb-Douglas Production Functions, Chicago 1965.

R. Pfouts, "The Theory of Cost and Production in the Multi-Product Firm," *Econometrica*, Oct. 1961, *29*, 650–68.

D. Saathoff and R. Kurtz, "Cost Per Day Comparisons Don't Do the Job," *Modern Hospital*, Oct. 1962, *99*, 14–16.

A. Walters, "Production and Cost Functions: An Econometric Survey," *Econometrica*, Jan.-Apr. 1963, *31*, 1–66.

A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Ass.*, June 1962, *57*, 348–68.

Blue Cross of Western Pennsylvania, *Comparative Financial and Statistical Information*, Pittsburgh 1966.

Discrimination and Income Differentials

By James Gwartney*

Income differentials between whites and nonwhites may result from several factors. The two populations differ in educational attainment, age distribution, geographic distribution, and rural-urban distribution; all of which are related to income and are, in some cases, determinants of it. When measuring employment discrimination, the relevant comparison is between individuals of similar productive capacity who differ only in color.[1]

This study seeks to break down the income differential between whites and nonwhites into two categories: (a) a differential resulting from differences in productivity factors not directly related to employment discrimination, and (b) a residual unaccounted for by differences in productivity factors and which may result largely from employment discrimination. Assuming homogenous preferences and standardizing the white and nonwhite populations for all differences in productivity factors that affect income, we could reasonably infer that any remaining income differential approximates income differences resulting from color discrimination in employment.

It should be noted that differences between whites and nonwhites in productivity factors may be the result of color discrimination in areas other than employment. The magnitude of the income differences resulting from differences in productivity factors will give some indication of the possible intensity of color discrimination in education and in other areas not related to employment discrimination.

If there were employment discrimination against nonwhites, they would be expected to receive lower incomes than whites of similar employability. If an employer had a preference for white labor relative to nonwhite labor purely because of color, he would be indifferent between white and nonwhite employees of similar employability only if the nonwhite employees were paid lower wages. White employees of similar employability would be preferred if the wage rate were identical for both whites and nonwhites. The greater the intensity of employment discrimination against nonwhites, the lower the nonwhite wage rate, and thus earnings, relative to that of whites of similar employability.

Employment discrimination could be present however, even if the money-wage rates were identical. The terms and conditions of the sale of labor by an employee include not only wages and related monetary compensation, but also working conditions and other nonpecuniary benefits. Individuals have some trade-off between preferred working conditions and additional money income. They would be willing to take less money income if the employment had offsetting nonpecuniary benefits. Since these factors result in a difference between money income and "total"

* The author is assistant professor of economics at Florida State University. He wishes to thank Kenneth McCaffree, John Floyd, J. B. Gillingham, Yoram Barzel, Al Holtmann, William Laird, and the referee for helpful suggestions on an earlier draft of this paper. The author also gratefully acknowledges the support of a grant from the Manpower Administration of the U.S. Department of Labor.

[1] The terminology of this paper includes locational factors such as regional and city-size distribution among "productivity factors." While locational factors are not often considered productivity factors, they do affect earnings capacity, and have the potential of contributing to income differentials between individuals and/or groups.

income, the total income of a white could be greater than for a nonwhite, even if their money incomes were identical, assuming the white received greater nonpecuniary job benefits.[2] Since all of the data of this paper are money income or earnings data, we can investigate only the consistency of those data with employment discrimination.[3] But money income is a sizeable component of total income, and where discrimination resulted in a significant total income differential between whites and nonwhites, we would expect to observe a money income differential as well.

In the following discussion, Section I estimates the importance of various factors in explaining the money income differential between white and nonwhite urban males, 25 years of age and over, in 1960.[4] Section II contains similar estimates using both mean and median earnings data for males in nonfarm occupations. Section III considers regional differences in earnings differentials between white and nonwhite males.

After adjustment for differences in education, scholastic achievement, age, region, and city size, the nonwhite median income is estimated between 81 and 87 percent of the white for urban males and 77 and 86 percent for males in nonfarm occupations. When mean income data are used, the estimated adjusted nonwhite/white income ratio is only slightly smaller than in the case of medians.

The adjusted nonwhite/white earnings ratio is estimated to be greater in the North than in the South. In the North for nonfarm occupations, the adjusted nonwhite/white earnings ratio is estimated between .83 and .88, compared to .68 and .74 for the South.

The results indicate that a large portion of the income differential between white and nonwhite urban males is the result of differences in quantity of education and scholastic achievement. Differences between whites and nonwhites in these two education-related factors are estimated to have accounted for nonwhite urban males receiving between 23 and 27 percent less income than white urban males in 1959. The size of these estimates indicate that unless differences between the two populations in these two factors can be reduced substantially, the median income of nonwhite males is unlikely to increase to more than 70 to 80 percent that of whites, even if employment discrimination is substantially reduced.

However, the unexplained differential in income between the two populations was estimated between 14 and 25 percent for nonfarm occupations and 13 to 19 percent for urban areas in the United States. This indicates that one-third to three-fifths of the nonwhite-white income differential remained unaccounted for after adjustment for the productivity factors of this paper and this residual may result largely from employment discrimination.

I. *Income Differences Between White and Nonwhite Urban Males in 1959*

The median income of nonwhite urban

[2] For example, a Negro might prefer to be a filling station attendant rather than a garbage collector. Strong employment discrimination might result in his becoming a garbage collector. His money income may be as great or greater, but his total income is less because he would have preferred employment as a filling station attendant.

[3] Wage and salary income, were it available, would be more appropriate than income data reflecting income from other assets in addition to earnings from labor. If there is proportionality between the two income measures, the relative incomes will be unaffected. On the assumption that relative income from other assets is a positive function of income level, the income data for whites may overstate their wage and salary income relative to nonwhites. However, if as seems likely, nonwhite income from labor is affected more by discrimination than their non-labor income, the bias will be in the opposite direction.

[4] M. Zeman estimated the significance of various factors in explaining income differences between whites and nonwhites in 1939. Due to data limitations, Zeman did not standarize for a number of factors considered in this paper.

males was 58.3 percent of white urban males in 1959. The income differential between white and nonwhite urban males is partially due to distributional differences in productivity factors.

This section seeks to estimate the income ratio of nonwhite to white urban males after correction for income differences resulting from differences in productivity factors. These factors include quantity of education, level of scholastic achievement, regional, age, and city-size distributions.

Three criteria were used in choosing productivity factors relevant to the explanation of income differentials between whites and nonwhites. First, only factors generally recognized as determinants of money income, or as closely correlated with income, were used in this study. Second, the factors chosen are *not* directly related to employment discrimination as such. For example, the median income of nonwhites in the United States is less than that of whites partially because of the overrepresentation of nonwhites in the low-income South—a situation which is not the result of employment discrimination.[5] By contrast, occupational distribution according to color is related to employment discrimination. Therefore, no adjustment was made for differences in occupational structure. Third, factors were either considered simultaneously, or chosen where the apparent relationship with other factors was one of independence. Thus even though low earnings are associated with youth, the factor of age is utilized in explaining the income differential only if the relatively low incomes are not the result of such other factors as size of cities or the regional distribution of population, with which both low earnings and age might also be correlated.

Standardization could either increase or decrease the magnitude of the observed income difference between whites and nonwhites. Factors with positive (negative) signs are those which tend to make the differential larger (smaller) than it would have been in the absence of those factors.

The income differential between whites and nonwhites is disaggregated by constructing indexes of *income* and *distributional* differences. *Distributional* differences in income result from differences in productivity factors between whites and nonwhites. *Income* differences result from differences in income between whites and nonwhites after adjustment for the productivity factors considered.

The Laspeyres index of *income differences* is the hypothetical ratio of the median income of nonwhite to white males, assuming both color groups were distributed among productivity categories as whites actually were. The Paasche index of income differences is a similar hypothetical ratio under the assumption that both color groups had the productivity distribution of nonwhites.[6] The index of income differences is essentially an estimate of the ratio of nonwhite to white income after adjustment for differences in the productivity factors considered.

The indexes of *distribution differences* represent the hypothetical ratios of nonwhite

[5] The intensity of discrimination in the South has, if anything, resulted in migration from the South which tends to decrease the magnitude of the regional distribution factor.

[6] The mathematical form of the index of income differences is:

$$\frac{\sum (Y_n \cdot D_w)}{\sum (Y_w \cdot D_w)} \text{ (Laspeyres) and, } \frac{\sum (Y_n \cdot D_n)}{\sum (Y_w \cdot D_n)} \text{ (Paasche)}$$

where,

Y is the median income of those with income within a productivity category (e.g. age, education or region) according to color,

D is the percent of population according to color with income within the productivity category, and

n and w are subscripts denoting nonwhite and white population groupings.

to white income if whites and nonwhites had the same income *within* productivity categories.[7] The index of *distribution differences* is essentially an estimate of income differences that result from differences in the distribution of productivity factors. The index of *income differences* multiplied by the index of distribution differences yields an index of "total" money income differences, which closely approximates the actual ratio of nonwhite to white income.[8]

The index of *income differences* will be less than unity if the median income of nonwhites is less than whites *within* productivity categories. The index of *income differences* is an estimate of the income ratio of nonwhites to whites if the two populations were similarly distributed among the adjusted productivity categories.

The index of *distribution differences* will be less than unity if nonwhites are overrepresented relative to whites in low income productivity categories. Such an index is an estimate of income differences resulting from differences in productivity capacity between the two populations. We proceed to adjust the data for the effects of differences in productivity factors that influence the ratio of nonwhite to white income.

1. *Quantity of Education Adjustment.*

White urban males have a greater quantity of education than nonwhites. Since income is positively related to education, whites would be expected to receive larger incomes than nonwhites on this basis. In isolating the "quantity of education" effect, the 1960 Census data indicating the median income of white and nonwhite urban males for each of eight different educational levels were utilized, (Table 223 of the *Final Report*, "Characteristics of the Population.") Both indexes of income and distribution differences were calculated. The Laspeyres and Paasche indexes of income differences were 67.0 and 69.9 percent, respectively, after correction for differences in the quantity of education distributions between white and nonwhite urban males.[9] The adjustment indicates that the income of nonwhite urban males was just over two-thirds the income of white urban males with similar quantity of education in 1960.

2. *Scholastic Achievement Adjustment.*

Whites and nonwhites differ not only in quantity of education, but also in what we shall term scholastic achievement level. National education tests show that nonwhites with the same number of years of formal schooling perform at a significantly lower level than whites. Some of the income differential between whites and nonwhites is the result of this lower achievement level.

The outlay of funds per pupil is often less for nonwhites than for whites, hence the lower achievement of nonwhite students may be the result of attending schools of inferior quality. Days of attendance vary among those with equal number of years of education. Nonwhites with equal quantity of education have fewer days of attendance than whites because of their overrepresentation in schools (largely in the South) with a six to eight month term

[7] The form of the index of distribution differences is:

$$\frac{\sum (Y_n \cdot D_n)}{\sum (Y_n \cdot D_w)} \text{ (Laspeyres) and, } \frac{\sum (Y_w \cdot D_n)}{\sum (Y_w \cdot D_w)} \text{ (Paasche)}$$

Since the distributional categories are in percent form, the estimates for those with income are not affected by differences in the size or participation rates of the two labor forces. One could construct an index of distribution differences assuming both color groupings had the median income of the total population within productivity categories. Since whites compose a large percentage of the total population, such an index would closely approximate the Paasche index.

[8] Also the inverse of the index of distribution differences multiplied by the ratio of total money income differences yields the index of income differences.

[9] The data and specific calculations for all estimates of this paper are contained in an Appendix that is available from the author on request.

and a seven year elementary degree requirement. In addition, even if nonwhites have a similar attendance record in schools of comparable quality, they may do less well than whites because cultural and environmental factors may hinder their taking advantage of educational programs reflecting white cultural standards, values, and traditions. Thus even when whites and nonwhites attend the same schools, it is possible, even probable, that a differential in achievement will result.

A recent study for the *U.S.* Office of Education by James Coleman, et al., estimates differences in scholastic achievement between whites and nonwhites in terms of quantity of education (years of schooling) for three different grade levels in metropolitan areas.[10] The estimates of scholastic achievement of the *Coleman Report* were derived from data based on 645,000 public school students' scores on the Educational Testing Service School and College Ability Test.[11] This test measures:

> . . . the skills which are among the most important in our society for getting a good job and moving up to a better one, and for full participation in an increasingly technical world. Consequently, a pupil's test results at the end of public school provide a good measure of the range of opportunities open to him as he finishes school—a wide range of choice of jobs or colleges if these skills are very high; a very narrow range that includes only the most menial jobs if these skills are very low. [p. 20]

The *Coleman Report* indicates that in metropolitan areas the scholastic achievement gap increases as the formal quantity of education increases. Within regions, nonwhites are behind whites by 1.4 to1.6 years at grade 6, by 2.2 to 2.5 years at grade 9, and by 2.9 to 3.6 years at grade 12. For whatever reason, these data indicate that a nonwhite with a sixth grade education is only as well qualified for employment as a white with 4.4 to 4.6 years of education and the employability differential increases at higher grade levels.[12]

Using the maximum scholastic achievement differential of the *Coleman Report* for grades 6, 9, 12, a maximum estimate of the scholastic achievement differential was projected under the assumptions that (a) the differential is linear, and (b) whites and nonwhites are of the same scholastic achievement level before entrance to school.[13] Under the same assumptions, plus the further assumption that the scholastic achievement differential for grade levels beyond twelve remained constant at the level of grade 12, a minimum estimate of the scholastic achievement differential was projected, using the minimum estimates of the scholastic achievement differential of the *Coleman Report*.

The income ratio of nonwhites to whites after adjustment for both quantity of education and scholastic achievement was estimated by comparing the actual income of nonwhite urban males in an education

[10] This study will subsequently be referred to as the *Coleman Report*.

[11] For additional comments on the *Coleman Report*, see Robert Dentler, Samuel Bowles and Henry Levin, Daniel Moynihan, and Robert Nichols. Much of the criticism of the *Coleman Report* concerns the causes of the estimated scholastic achievement differential, rather than its magnitude.

[12] Because the *Coleman Report* did not disaggregate data for nonwhites as a group, these differentials are for Negroes. However, except for Oriental Americans, the differential between whites and other nonwhite groups was similar to the Negro differential. In addition, since 95 percent of all nonwhites are Negroes, one would expect the differential between whites and Negroes to be almost identical to the differential between whites and nonwhites.

[13] Recent studies indicate that nonwhites are, in fact, already behind whites upon entrance to school. The results of the *Coleman Report* (p. 20) indicated scores of minority pupils are as much as one standard deviation below the majority pupils' scores in the first grade. No adjustment was attempted for this differential because comparative income data for whites with no education and low achievement would not be available. Failure to make adjustment for this factor will tend to understate the ratio of nonwhite to white income after correction for the scholastic achievement factor.

cell with the income of white urban males equal in achievement level with the nonwhites of the education cell.[14] The results indicate that in 1960 the income of white urban males was between 12.2 (Paasche) and 18.1 (Laspeyres) percent greater than nonwhite urban males because of the lower achievement level of nonwhites.[15] Cumulatively, the adjusted nonwhite/white income ratio was estimated between 82.1 and 85.1 (Table 1).[16]

[14] The form of the adjusted index of income differences between whites and nonwhites is:

$$\frac{\Sigma(Y_n \cdot E_w)}{\Sigma(Y'_w \cdot E_w)} \text{ (Laspeyres) and, } \frac{\Sigma(Y_n \cdot E_n)}{\Sigma(Y''_w \cdot E_n)} \text{ (Paasche)}$$

where, Y_n is the median income of nonwhite urban males with income in 1959 for each of the eight different quantity of education cells,

Y'_w and Y''_w are respectively, maximum and minimum estimates of the incomes of white urban males of equal achievement level with the nonwhites of the education cell, and

E_w and E_n represent the percent of white and nonwhite urban males within each education cell with income in 1959.

It is possible, using both the maximum and minimum estimates of income adjusted for scholastic achievement differences within an education cell, to make two estimates of the Laspeyres and Paasche indexes of income differences.

[15] The data used are from 1960 Census of Population, "Characteristics of the Population" Table 223. The income data for white males of equal achievement level with nonwhites of a given education cell were derived by interpolation, using the achievement level differentials of the *Coleman Report* and the assumptions stated in the text of this paper. The data of the *Coleman Report* were not disaggregated according to sex, thus the achievement level differences were assumed to be the same for males as for the total.

[16] The data of the *Coleman Report* were obtained in September 1965. Since those covered by the 1959 income data completed their education anywhere from several years to a half century or more before 1965, the differential in level of scholastic achievement is assumed constant over time. Increased expenditures on nonwhite education relative to white and Supreme Court decisions on desegregation suggest a possible narrowing of achievement differentials in recent years inasmuch as they are related to differences in the quality of schools. If there has been a narrowing of the achievement level differentials over time, the estimate of the adjusted nonwhite/white income ratio is biased downward. However, the little evidence existing on comparative test scores over time does not confirm a narrowing of achievement level differentials, but indicates there has been little change (see John Miner and A. M. Shuey). In any case, there would not seem to be any a priori reason to suggest a smaller scholastic achievement differential for the adult population in 1959 than was the case for the school age population of 1965.

3. *State Distribution.*

Nonwhites are overrepresented in southern states. This reduces their income relative to whites because (a) incomes of both whites and nonwhites are lower in southern states than northern states and (b) the income differential is greater in southern states.

The median income and education distributions according to color are available for urban areas by state from the 1960 Census, *Detailed Characteristics* (Tables 47 and 138). This provides the information to calculate the indexes of income and distribution differences, simultaneously adjusting for state distribution and quantity of education difference.[17] The marginal effect of the state distribution factor can be estimated by comparing the indexes of income differences after adjustment for the state distribution and quantity of education factors with the indexes when adjustment was made for quantity of education only. Thus, double counting of regional income differences that are really the result of educational differences between regions can be avoided.

After adjustment for both quantity of education and state distribution, nonwhite urban males are estimated to receive incomes between 72.3 (Paasche) and 72.9 (Laspeyres) percent as large as incomes of white urban males. Comparing these indexes with the indexes when only quantity

[17] This was done for all states with a nonwhite population of 10,000 or more. These states include more than 98 percent of the nonwhite population in the United States. Due to data limitations, the index of income differences for northern states was estimated from the index of distribution differences assuming both whites and nonwhites received the income of the total population. The Laspeyres (Paasche) index of income differences for the North was derived by weighting the state indexes according to the percentage of all whites (nonwhites) that resided in that state.

TABLE 1—INCOME OF NONWHITES AS A PERCENTAGE OF WHITES, ADJUSTED FOR VARIOUS DETERMINANTS OF INCOME DIFFERENTIALS BETWEEN WHITE AND NONWHITE URBAN MALES FOR THE UNITED STATES IN 1959

	Index of Income Differences[a]		*Marginal Effect of Factor*	
	Laspeyres	Paasche	Laspeyres	Paasche
Unadjusted Income Ratio (Nonwhite/White)	58.3	58.3	—	—
Explanatory Factors				
A. Quantity of Education	67.0	69.9	8.7	11.6
B. Scholastic Achievement	85.1	82.1	18.1	12.2
C. State Distribution	91.0	84.5	5.9	2.4
D. City Size	89.3	83.3	− 1.7	− 1.2
E. Age Distribution	86.5	80.9	− 2.8	− 2.4

[a] Since both a maximum and minimum scholastic achievement differential was used from the *Coleman Report*, it is possible to calculate two Paasche and Laspeyres indexes of income differences when adjusting for this factor. The above estimates of the scholastic achievement factor "bracket" the other two estimates that could be calculated. Therefore, the procedure yields both a maximum and minimum estimate of the nonwhite/white income ratio adjusted for the five productivity factors. The same procedure was also followed for the estimates of Tables 2 and 3.

Source: Estimates are derived from *U.S.* Census data, "Characteristics of the Population," Tables 219 and 223; *Subject Report, Size of Place*, Tables 2, 3, 4, and 5; *Detailed Characteristics*, Tables 47 and 138.

of education was adjusted for, the data indicate the income of nonwhite urban males increased between 2.4 (Paasche) and 5.9 (Laspeyres) percent relative to that of white urban males as the result of the marginal effect of the state distribution adjustment. The Laspeyres adjustment, which assumes that both populations had the distributional characteristics of whites, is greater than the Paasche estimate because of a smaller income differential in the North—the area where whites are relatively overrepresented. Cumulating the adjustments for (a) quantity of education, (b) scholastic achievement, and (c) state distribution, the income of nonwhite urban males is estimated between 84.5 and 91.0 percent of white urban males (Table 1).

4. *City-Size Distribution.*

Nonwhites are overrepresented relative to whites in the larger cities. Median income differs among cities of different size, tending to increase as the size of the city increases. Nonwhite overrepresentation in the larger cities would therefore result in the income of nonwhites being greater than if they were distributed among cities in the same manner as whites.

Data from the 1960 Census gives the median income of the total population and of nonwhites according to size of city by region, as well as the percentage of the white and nonwhite population residing in cities of each size, (see Tables 2, 3, 4, and 5, *Subject Report, Size of Place*). Thus, both a Laspeyres and Paasche index of distribution differences can be constructed to estimate the amount by which nonwhite income exceeds white income because of the overrepresentation of nonwhites in large cities. The indexes are calculated by region, thus minimizing interaction with the state distribution factor.

The results of the adjustment indicate that the incomes of nonwhite urban males were between 1.2 (Paasche) and 1.7 (Laspeyres) percent higher relative to white males because of this factor. Adjust-

ing for the city-size factor increases the white/nonwhite income differential—hence the index of distribution differences is greater than one and the sign of the factor is negative. The Laspeyres index, which assumes both populations were distributed among cities in the same manner as whites, is greater than the Paasche index because the city-size differential is greater in the North. After the adjustment for city size in addition to the three previous factors, the nonwhite/white income ratio for urban males is estimated between 83.3 and 89.3 percent.

5. *Age Distribution.*

Nonwhite urban males were overrepresented in the prime earning age categories in the United States in 1959. While whites are overrepresented in older age categories, particularly those over 65, nonwhites are overrepresented among the young. However, the larger percentage of the white urban male population among those past the years of prime earnings more than offsets the youth of the nonwhite population.[18]

Data are available to calculate both the Laspeyres and Paasche indexes of distribution differences ("Characteristics of the Population," Table 219). The results indicate that if the age distribution of white urban males in 1959 had been the same as that of nonwhite urban males, the income of nonwhites relative to whites would have been between 2.4 (Paasche) and 2.8 (Laspeyres) percentage points lower than the observed ratio (i.e. negative in explaining the income differential).

Summary of Estimates for Urban Males.

Table 1 consolidates the separate estimates of the magnitude of various factors upon the earnings differentials between white and nonwhite urban males. While the unadjusted income of nonwhites was only 58.3 percent as great as whites, the income of nonwhite urban males is estimated between 81 and 87 percent of the white income after adjustment for the five factors of Table 1. An income differential of between 13 and 19 percent remains unexplained.

II. *Earnings Differentials Between Whites and Nonwhites in Nonfarm Occupations in 1959*

The estimates of Section I might be in error because it was not always possible to adjust for all factors simultaneously.[19] In addition, only median income data were used, and mean income estimates would be useful for purposes of comparison.

Mean and median earnings data for males according to color, age, education, and region were obtained for nonfarm occupations from 1960 Census data, *Subject Report, Occupation by Earnings and Education.* This data will allow for the simultaneous adjustment for all factors of Section I, except city-size.

The data used in this section differ from that of Section I in four respects. First, the data used are for earnings rather than income. Earnings data, unlike the income data used in Section I, include only wage and salary and self-employment income.[20] Second, they are for males with income between 25 and 64 years of age in nonfarm occupations, rather than males 25 and over in urban areas. Third, the quantity of education data used in Section I was more refined, having eight education categories

[18] Since the data cover those over age 25 with income, the youth of the nonwhite population is less important than if the data were for all with income. Census data used contained six age classes: 25–34, 35–44, 45–54, 55–64, 65–74, and 75 years of age and over.

[19] Obvious interaction between such factors as region and quantity of education was avoided by estimating their effect simultaneously.

[20] Income data includes, in addition to earnings, income from Social Security, pensions, Veterans payments, rents, interest or dividends, unemployment insurance and welfare payments.

TABLE 2—EARNINGS OF NONWHITES AS A PERCENTAGE OF WHITES, ADJUSTED FOR VARIOUS DETERMINANTS OF EARNINGS DIFFERENTIALS FOR NONFARM OCCUPATIONS FOR THE UNITED STATES IN 1959.

	Index of Income Differences		*Index of Distribution Differences*	
	Median	Mean	Median	Mean
Unadjusted Income Ratio (Nonwhite/White)	60.1	54.6	—	—
Explanatory Factors				
A. Region				
Laspeyres	68.1	59.9	88.3	91.2
Paasche	63.7	55.6	94.3	96.8
B. Region-Age				
Laspeyres	66.5	58.9	91.8	92.7
Paasche	62.1	55.6	98.2	98.2
C. Region-Education				
Laspeyres	72.0	64.4	84.4	84.8
Paasche	71.0	64.8	85.7	84.3
D. Region-Age-Education				
Laspeyres	71.2	65.4	84.3	83.5
Paasche	70.8	67.1	84.8	81.3
E. Region-Age-Education-Scholastic Achievement				
Laspeyres	85.6	83.6	70.2	65.3
Paasche	77.2	75.2	77.8	72.6

Source: The data used in calculating the estimates of this table were derived from *U.S.* Census data, *Subject Report, Occupation by Earnings and Education*, Tables 1, 2 and 3.

rather than only six.[21] Fourth, estimates of the previous section were based on adjustment for state distribution of the populations, while data of this section will allow for only a North–South regional adjustment.

The adjustment techniques used in this section are identical to those of Section I: the Laspeyres (Paasche) index calculated assuming both whites and nonwhites are distributed among productivity categories in the same manner as whites (nonwhites). Table 2 summarizes the results.

[21] The data used in this section classifies those with 0 to 7 years of education as one group, rather than the 0, 1–4, 5–7 year classes previously used. This is important since a large percentage of the non-white population is in this class. The impact of using the broader educational class will result in an under-estimation of (a) the quantity of education adjustment and (b) the adjusted nonwhite/white income ratio.

After simultaneous adjustment for the four factors considered in Table 2, the median earnings of nonwhite males in nonfarm occupations is estimated between 77 and 86 percent of the earnings of whites. For means, the adjusted earnings ratio is estimated between 75 and 85 percent, 2 percent less than for the medians.

The estimates are consistent with those of Section I. Adjustment for age had little effect after the region-education adjustment was considered. The failure of age to have the negative impact estimated in Section I results from the exclusion of those over 65 from the data used in this section. Since whites are highly overrepresented in the over-65 group, their inclusion increased the observed nonwhite/white income ratio and resulted in the negative contribution of age found in Section I. Table 2 estimates

that adjustment for region-education differences would increase the income of nonwhites by between 10 and 12 percent relative to whites. This is between 2 and 4 percent less than the estimates of Section I. Since this adjustment was made by exactly the same techniques in both cases, the difference between the two estimates must be attributable to the use of broader educational and regional classes in this section.

The median earnings estimate of the scholastic achievement factor of Table 2 is slightly less than the estimate of Section I. When mean earnings data are used, the estimated importance of the scholastic achievement factor increases. The magnitude of the scholastic achievement factor is greater for means than medians because the rate of change in the education-income relationship is greater for means than medians. Thus even though the unadjusted differential was 5 percent greater for means than medians, after adjustment for the four factors of Table 2 the difference was reduced to only 2 percent.

The estimates of this section are not adjusted for city-size, but this deficiency, resulting in an over estimate of the nonwhite /white ratio, is almost surely offset by the inability to use the more refined educational and regional categories of Section I.[23] The data indicate that after adjustment for the four factors of Table 2, an income differential between 14 and 23 percent remained unexplained when median earnings data were used. In the case of mean earnings data the unexplained differential is estimated between 16 and 25 percent—slightly greater than for medians. Thus, adjustment for the factors of this section "explained" 40 to 65 percent of the total income differential between whites and nonwhites in 1959. The probable effect of using mean rather than median income data in the analysis underlying Table 1 can be inferred from the index of income differences between means and medians of this section.

The estimates of the adjusted ratio of nonwhite/white income of Section I are consistent with the estimates of this section. Due to the factors referred to in footnote 22 we might have expected these estimates to be slightly less. Considering the expected downward bias, the results of this section reenforce our estimates of Section I.

III. *Regional Differences in the Nonwhite/White Earnings Ratio for Nonfarm Occupations*

Previously we indicated the importance of the regional (state) distribution factor on the nonwhite/white relative income for the United States. Now, we investigate *regional* differences in these income differentials. The South is generally believed to be more inclined toward employment discrimination than the North. Census data indicate the median income of nonwhite males in 1959 was 47 percent that of white males in the South and 73 percent in the North. The relative income of nonwhites to whites in the South was only 65 percent of the same ratio in the North.

At least three factors unrelated to employment discrimination, although not necessarily unrelated to other forms of color discrimination, result in making the income data of the South appear very unfavorable relative to the North. First, the quantity of education of nonwhites relative to whites is much less favorable in the South. Second, the nonwhite population in the North is significantly overrepresented in urban areas. Third, the nonwhite

[23] Comparison of the estimates of the region-education adjustment of this section with the state distribution-education adjustment of Section I indicate the more refined classifications of Section I increase the contribution of these two factors by 2 to 4 percent. The city-size factor of Section I was 1.2 to 1.7 and negative in sign. Use of the broader educational classes and failure to adjust for the city-size factor will result in the estimates of this section being 1 to 3 percent less than the estimates of Section I.

TABLE 3—EARNINGS OF NONWHITES AS A PERCENTAGE OF WHITES ADJUSTED FOR VARIOUS DETERMINANTS OF EARNINGS DIFFERENTIALS BY REGION FOR NONFARM OCCUPATIONS IN 1959

	Index of Median Income Difference		*Index of Mean Income Difference*	
	North	South	North	South
Unadjusted Income Ration (Nonwhite/White)	72.7	52.1	63.7	47.1
Explanatory Factors				
A. Age				
Laspeyres	72.8	52.2	63.4	47.2
Paasche	72.4	52.2	63.8	47.2
B. Education				
Laspeyres	65.1	58.9	68.0	53.9
Paasche	67.1	61.9	70.8	58.2
C. Age-Education				
Laspeyres	75.8	58.6	68.4	54.6
Paasche	79.5	62.1	72.2	59.7
D. Age-Education-Scholastic Achievement				
Laspeyres	87.9	73.9	86.0	72.5
Paasche	83.6	67.9	81.3	66.5

Source: See Table 2.

population in the North is not only overrepresented in urban areas, but specifically in large urban areas with high money income.

Table 3 disaggregates the estimates of Section II by region. After simultaneous adjustment for the four factors of Table 3, the median earnings ratio of nonwhites to whites is estimated between .83 and .88 in the North and .68 and .74 in the South. Similar differences in the adjusted mean earnings ratio exist between the two regions. The adjusted relative earnings estimates of the nonwhite/white income ratio for nonfarm occupations (Table 3) for the South are 82 and 83 percent of similar estimates for the North.[23] Standardization for differences in the productivity factors between the two populations reduces, but does not eliminate, the North–South differential for relative income of nonwhites to whites.

Estimates similar to those of Table 3, using the same data can be derived by regression analysis. When the logarithm of income is the dependent variable and a dummy variable (white 1 and nonwhite 0) is used for color, the regression coefficient of color is the antilogarithm of the white/nonwhite income ratio.[24] Table 4 presents

[23] The North-South differential is slightly overstated because of the greater city-size contribution in the North. The city-size factor in the North is estimated between 1.5 and 1.8—in the South 0.9 and 1.6. In both cases it is negative, indicating the actual nonwhite/white income ratio is overstated because of the failure to adjust for this factor. In addition, it is usually argued that less money per head is spent on the education of southern Negroes as compared to northern Negroes, resulting in an inferior quality of education for the former. However, the *Coleman Report* indicated the white/nonwhite scholastic achievement differential for a given grade level was nearly as great in the North as the South. The existence of a greater scholastic achievement differential in the South would decrease the adjusted North-South differential.

[24] The regression coefficient for color is $\log Y^w - \log Y^n$, or $\log (Y^w/Y^n)$. The nonwhite/white earnings ratio (Y^n/Y^w) contained in Table 4 is the inverse of the antilogarithm of the color coefficients.

TABLE 4—REGRESSION EQUATIONS ESTIMATING THE NONWHITE/WHITE EARNINGS (Y^n/Y^w) RATIO AFTER ADJUSTMENT FOR VARIOUS PRODUCTIVITY FACTORS
(t-ratios in parentheses)[b]

Logarithm of Median Earnings[a]	b_0[c]	X_1	X_2	X_3	X_4	X_5	X_6	R^2 (df)	Y^n/Y^w
1. North	3.461	.060 (4.78)	.057 (4.52)	.034 (2.71)	.022 (14.99)	−.042 (4.10)	.0326 (1.32)	.791 (41)	.928
2. North	3.430	.060 (4.22)	.057 (4.04)	.034 (2.42)	.021 (11.99)	−.017 (2.30)	.0760 (3.10)	.776 (41)	.839
3. South	3.232	.069 (4.56)	.063 (4.23)	.020 (1.34)	.031 (17.51)	−.056 (4.58)	.1072 (3.60)	.817 (41)	.781
4. South	3.190	.069 (3.99)	.063 (3.70)	.020 (1.18)	.029 (13.78)	−.023 (2.56)	.1645 (5.43)	.806 (41)	.684

X_1=dummy—35–44 age group
X_2=dummy—45–54 age group
X_3=dummy—55–64 age group
X_4=years of education of those with income in the age education cell
X_5=scholastic achievement differential between whites and nonwhites
X_6=dummy for color (white 1 and nonwhite 0)

[a] Odd (even) numbered equations are for data containing the maximum (minimum) estimate of the scholastic achievement differential from the *Coleman Report*.

[b] Critical values for t are approximately ± 2.42 for 99 percent, ± 1.68 for 95 percent, ± 1.30 for 90 percent, ± 1.05 for 85 percent.

[c] The coefficient b_0 is an estimate of the median earnings of nonwhite males in the 25–34 age group. The interpretation of the last column (Y^n/Y^w) is given in fn 24.

such regression equations by region, using median earnings as the dependent variable. In addition to the color variable, adjustment was made for differences in education (X_4), age (X_1, X_2, X_3), and scholastic achievement (X_5). The regression was run using both the maximum and minimum estimates of the scholastic achievement differential between whites and nonwhites.

Nonwhite earnings are estimated between 83.9 and 92.8 percent of the white earnings in the North, and 68.4 and 78.1 percent in the South. The color variable is significant at the 0.90 level in all cases. The results also indicate that even after correction for age, education, and scholastic achievement differences, the adjusted nonwhite/white earnings ratio is greater in the North than the South.

The estimates of the nonwhite/white earnings ratio presented in Table 4 are just slightly greater than those of Table 3. The small difference resulted because, after adjustment for the scholastic achievement factor, the nonwhite/white income ratio increases as quantity of education increases, i.e. the differential is less for higher education cells. When using regression analysis, all observations receive equal "weight." However, when using the method of Table 3, observations in the highest education cells receive less weight because of the smaller number of individuals in those cells. Thus, there was a small difference between the estimates.

The estimates of Table 4 reenforce our previous estimates that after adjustment for the productivity factors of this paper, an unexplained income differential between white and nonwhite males still remains and the unexplained differential is larger in the South than the North.

REFERENCES

G. Becker, *The Economics of Discrimination.* Chicago 1957.

S. Bowles and H. M. Levin, "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *J. Hum. Resources*, spring 1968, *3*, 1–24.

J. S. Coleman, et al., *Equality of Educational Opportunity*. Washington 1966.

R. A. Dentler, "Equality of Education Opportunity: A Special Review," *Urban Rev.*, Dec. 1966, *1*, 27–29.

J. B. Miner, *Intelligence in the United States*. New York 1957.

D. P. Moynihan, "The Crisis of Confidence," presented to the Subcommittee on Executive Reorganization of the U.S. Senate Committee of Government Operations, Dec. 13, 1966.

R. C. Nichols, "Schools and the Disadvantaged," *Science*, Dec. 9, 1966, *154*, 1312–14.

A. M. Shuey, *The Testing of Negro Intelligence*. New York 1966.

M. Zeman, "A Quantitative Analysis of White-Nonwhite Income Differential in the United States in 1939," unpublished doctoral dissertation, Univ. Chicago 1955.

U.S. Census of Population: 1960, *U.S. Summary, Final Report*, "Characteristics of the Population."

———, *Subject Report, Size of Place*, PC (3)-1B.

———, *Subject Report, Occupation by Earnings and Education*, PC (2)-7B.

———, *Detailed Characteristics, Final Report*, PC(1) Series.

Schooling and Earnings of Low Achievers

By W. Lee Hansen, Burton A. Weisbrod, and William J. Scanlon*

Why do earnings differ among people? Much has been made in recent years of the importance of years of schooling as a determinant of earnings. But whatever is true of the contribution of schooling *in general*, its contribution for persons at the lower end of the achievement spectrum may be quite different than at the top. If so, the importance of schooling may have been inadequately emphasized for some groups and overemphasized for others by focusing on average benefits.

The objective of this paper is to estimate the extent to which schooling, as contrasted with "learning" and job training, is an important determinant of earnings of low achievers, and to estimate the importance of a number of personal and environmental factors influencing their earnings levels. It seems especially important to know more about the impact of education on lower achievers, given the propensity of many to view education as the principal and most effective device for raising earnings, particularly among the poor and the disadvantaged. Such a view lies behind campaigns to curb high school dropouts and other proposals to provide more and better education for the disadvantaged.[1]

It is well known that level of educational attainment (*LEA*) is positively correlated with level of earnings; for males and females, for whites and for nonwhites, and for virtually all age groups. At the same time, it is recognized as erroneous to attribute all of the observed differentials in earnings associated with education to differences in years of schooling. For it is also of importance to know how much people learned while in school, as a reflection of systematic differences in the ability, motivation, and family backgrounds of students, and in the quality of the schooling they receive. Support for the view that these and other factors, many unrelated to education, explain at least some portion of the earnings differentials commonly associated with differences in *LEA* rests in part on the observation that although mean incomes do differ significantly among educational groups, there is still extensive overlapping of the distributions around the means.[2] However, the problem has been an inability to isolate the quantitative importance of the school-related variables with any great degree of precision, despite some noteworthy attempts.[3]

What are the determinants of a person's earnings? We can quickly assemble a catalog: (1) physical conditions, including one's general state of health and presence of any disability; (2) learning and experience, determined by the quantity and quality of formal education and job training, innate ability, and the collection of experiences which contribute to a person's

* University of Wisconsin. This is a portion of a larger study dealing with the relationships among education, ability, and income, supported by the Ford Foundation. We wish to acknowledge the financial support provided by the Ford Foundation and the Institute for Research on Poverty, University of Wisconsin. The basic data were generously made available to us by the *U.S.* Department of Labor, Office of Manpower, Automation and Training. We are grateful to a number of people for their comments, particularly Arthur S. Goldberger, Samuel Bowles, and Thomas Ribich.

[1] For some evidence of the relatively low effectiveness of such programs in terms of monetary benefits relative to costs, see B. A. Weisbrod (1965, pp. 117–49) and T. Ribich.

[2] See *U.S.* Bureau of the Census.

[3] See: G. S. Becker, D. S. Bridgman, D. Wolfle and V. G. Smith, E. G. Denison, S. J. Hunt, and B. A. Weisbrod and P. Karpoff.

knowledge and skills; (3) psychological characteristics, among them work-leisure preferences, motivation, and the ability to communicate and cooperate in work situations; (4) family environment, reflecting informal learning in the home, and also the contacts and opportunities that the family can provide; (5) job access, which includes economic opportunities in the area as well as the degree of racial and other types of discrimination. This listing is by no means complete, nor are the classes indicated mutually exclusive, but the number and complexity of the forces likely to affect earnings are abundantly clear.

We will focus the bulk of our attention on factor (2), with some additional attention given to (4) and (5). We make no pretense in this paper to having a fully specified empirical model. However, we believe that we have been able to develop useful proxies for a number of these variables and to learn something about their quantitative importance for low achievers.

I. *Determinants of Earnings*

In this section we set forth a series of four alternative single-equation regression models of the determinants of individuals' earnings. Throughout we pay particular attention to the role of schooling, noting how its apparent importance varies as additional explanatory variables are considered. These models are developed in the light of the data available for our analysis, and, thus, even the most elaborate among them takes into account only a handful of the many variables that we believe to be relevant.

We begin with Model I in which earnings—and, hence differentials in earnings—within age-sex-race groups are attributed solely to level of schooling. While such a model is naively simple, it does embody the approach implicitly reflected in the frequently heard statements about the financial payoff from education in today's world, statements based largely on crude census data classifying earners by size of earnings and level of educational attainment plus age, sex, and race.

Analyses of the relationship between schooling and earnings have long recognized the incompleteness of an approach that emphasizes only, or even primarily, level or quantity of schooling. The quality of schooling, the motivation and ability of the student, and, in turn, what the student has learned are surely important. Indeed, one might hypothesize that it is *only* what a student has learned, both in school and elsewhere, and not the amount of time spent in school, that influences earnings. Accordingly, we propose the addition of a variable that captures such learning (Model II).

Besides schooling and learning, job training is likely to be an important means of acquiring knowledge and skills useful for enhancing earnings. Indeed, a variety of studies have demonstrated the substantial payoff of job training.[4] Thus, we include in Model III variables reflecting the quantity and quality of training.

At this point, having considered variables related to the learning process, we add a group of personal and environmental variables to mirror individual and family characteristics and regional economic opportunities (Model IV). Individual and family circumstances can affect the development of personality, mental capability, motivation, and through them affect earnings. At the same time they will affect earnings as a result of the contacts and the opportunities which family position can command. Similarly, economic opportunities vary among regions, reflecting differential prices and unemployment rates, differences in the mixes of skills demanded and supplied, access to knowledge about

[4] See J. Mincer, D. A. Page, G. G. Somers and E. W. Stromsdorfer and M. E. Borus.

job opportunities, and the like.[5] Since the ways in which these variables operate on earnings are sometimes less clear than with the education and learning variables, our empirical efforts represent a crude foray into this realm.

II. *Empirical Application*

In this section we describe the sample data, list the proxies for the classes of variables discussed above in connection with each of the models and explain the rationale for selecting the particular proxies used. Finally, we present and analyze the findings.

The Data

In November of 1963, the President's Task Force on Manpower Conservation interviewed a national sample of approximately 2,400 men, aged 17–25, who were rejected for military service because of failure to pass the Armed Forces Qualification Test (*AFQT*). The Task Force objective was to learn more about the characteristics of these mental rejectees as a guide to policy for cutting the rejection rate. The data obtained also serve us in our effort to study determinants of earnings. The sample data are unusual in that they include for each man, information on his earnings, schooling, *AFQT* score, training, and a variety of other factors, at least some of which are thought to influence levels of earnings. Of particular note are the *AFQT* scores which are at or below the 30th percentile. These percentile scores are not based on the current group of test takers, but rather on a standardized group—all men in military service in 1944.[6]

The fact that the sample consists entirely of low achievers (as measured by *AFQT*) makes it clearly a biased sample of the population of all young men, and so we cannot generalize our findings beyond the range of our observations—that is, to higher achievers. However, the size of the rejectee group is itself sufficient to warrant our interest, since it comprises 12 percent or more of all males aged 17–25,[7] and low achievers presumably constitute much of the core of the poverty problem.

The Variables

The dependent variable, annual earnings, Y, indicates total money income received in the year 1962, after deduction of transfer payments. While there are always problems in knowing exactly what is actually included in reported income, we are willing to accept the measure as a reasonably good proxy for market earnings.

Schooling, X_1, is measured simply as years of educational attainment, without regard for type (academic versus vocational) or quality of schooling.

As a broad measure of learning (either in or out of school), reflecting student ability, motivation, school quality, and home environment, we used the percentile score on the *AFQT*, X_2. The test is rather comprehensive, covering word knowledge, arithmetic, mechanical understanding, and ability to distinguish forms and patterns.

> The examinee's score on the test depends on several factors: on the level of his educational attainment; on the quality of his education (quality of the school facilities); and on the knowledge he gained from his educational training or otherwise, in and outside of school. These are interrelated factors, which obviously vary with the youth's socioeconomic and cultural environment, in addition to his innate ability to learn—commonly understood as I.Q.[8]

Our proxy for job training, X_3 is far from ideal, since it is a dummy variable indicating only whether or not an individual received training outside of school.

[5] See V. R. Fuchs.

[6] See B. D. Karpinos (1966).

[7] President's Task Force on Manpower Conservation, p. 11.

[8] See Karpinos, 1965, p. 6.

Neither the intensity, the duration, nor the type or quality of the training is known. This variable must also be regarded as a weak proxy for economic environment since the acquisition of training depends upon its availability and expected payoff.

Throughout the analysis we control for age, X_4, recognizing that it reflects a conglomeration of factors affecting earnings. For example, it measures the acquisition of experience from a variety of sources—schooling, job training, and employment. It also reflects the depreciation and obsolescence of learning and training. Finally, it captures some of the variation in motivation that is correlated with family responsibility. Because we will be concerned in our sample with young men aged 17–25, this family-responsibility element seems especially important; within this age span there is greatly reduced dependence upon parents as more and more individual responsibility is assumed.

Another control that we employ throughout is color, X_5. Differences in mean earnings between whites and nonwhites have often been noted, along with numerous possible causes. Although our data offer no opportunity to isolate the causes of these differences—job discrimination and low-quality schooling are especially suspected—the inclusion of the color variable will improve the estimation of the effects of our other variables.

Turning to the personal characteristics variables, we assume that being married indicates increased financial responsibility. and thus current marital status, X_6, reflects one aspect of motivation. A positive sign is expected for the coefficient of this variable. The variables denoting divorce of parents, X_7, and family size during childhood, X_8, are included to describe the individual's family circumstances as a youth. While the role of the family in affecting earnings is almost impossible to quantify, these variables should capture at least some of the disadvantages accompanying the divorce of parents and of life in a large family. Negative signs are anticipated for X_7 and X_8.[9]

The region variable, X_9, serves as a proxy for the economic environment. This variable will indicate the increment to income from living in a region outside of the South, and, accordingly, a positive sign is expected; it should capture some of the effects of regional price differences and employment opportunities.

Before we proceed to the findings, we need to note that there are two dimensions to the definition of low achiever. In addition to *how* to measure low achiever—for which we have used *AFQT* score—there is also the issue of *when* to measure it. Low achievers could be defined in terms of achievements at various points in time, depending on the type of question one sought to answer. For example, one might be interested in the question: What is the effect on income of different amounts of subsequent schooling for youngsters who are judged to be 'low achievers' at, say, age 10? In this case the appropriate population from which to sample is that of low achievers at age 10.

An alternative question is: "What has been the effect on income of differential amounts of schooling for men aged 17–25, who, following completion of their schooling, were judged to be low achievers?" To answer this requires sampling from a population of men who are low achievers at age 17–25. This is the nature of the sample in this study.

Of course, the two populations could be essentially the same. This would be the case if all low-achieving youngsters remained low achievers as young adults, and if no additional young people moved

[9] These personal characteristics variables are a set of dummies: X_6 is 1 if married; 0 if not married. X_7 is 1 if parents divorced; 0 if not. X_8 is 1 if 5 or more brothers and sisters; 0 if less than 5.

TABLE 1—REGRESSION RESULTS

All Men ($N=2403$)

Model	Constant	Education X_1	*AFQT* X_2	Training X_3	Age X_4	Color X_5	Marital Status X_6	Divorce of Parents X_7	Family Size X_8	Non-South X_9	R^2	Standard Error of Estimated Earnings
I	195.8 (135.3)	61.5* (12.1)	***	***	191.1* (15.2)	607.5* (58.4)	***	***	***	***	0.107	1399.57
II	224.5 (134.5)	30.3** (13.2)	26.6* (4.6)	***	201.7* (15.2)	537.0* (59.3)	***	***	***	***	0.120	1390.02
III	253.6 (134.4)	25.7 (13.2)	25.9* (4.5)	326.4* (90.6)	199.0* (15.2)	518.6* (59.4)	***	***	***	***	0.124	1386.56
III_a	227.4 (135.2)	55.9* (12.2)	***	345.9* (91.2)	188.6* (15.2)	586.3* (58.5)	***	***	***	***	0.113	1395.68
IV	252.1 (146.0)	20.3 (13.2)	23.7* (4.5)	291.3* (89.3)	184.0* (15.2)	361.7* (63.1)	458.7* (80.5)	−60.1 (61.9)	−40.4 (58.0)	442.0* (60.7)	0.155	1363.18

* Indicates that the variable is significant at the 0.01 level.
** Indicates that the variable is significant at the 0.05 level.
*** Indicates that the variable did not enter that particular model.
Standard errors are in parentheses.

into the low-achiever group between ages 10 and 17–25. But this is not likely, at least not in the extreme. Thus, while the present study does shed light on the second question, it also helps to answer the first, to the extent that the low-achiever populations among children and among men 17–25 are essentially the same. But insofar as some youthful low achievers might have escaped this status by age 17–25, and did so as a result of subsequent schooling beyond age 10, the estimates presented here understate, and therefore provide a lower-bound estimate of, the effect of schooling on incomes of all *youthful* low achievers.

The Findings

Each of the models described in Section I has been estimated by least squares, in the linear form $Y=a_0+a_1X_1+\ldots+a_nX_n$. The results for the entire sample of 2,403 men are presented in Table 1.[10] It should be recalled that all of these results apply to low-achiever males, as described earlier. It should be noted that because all men in the sample are low achievers does not preclude, or indeed, even restrict, the possibility of finding that additional schooling or additional learning bring financial benefits. There is considerable variation, within the sample, of incomes, years of schooling, *AFQT*, and, indeed, of all other variables considered in this paper. For example, years of schooling has: 578 observations with fewer than 8 years; 443 with 8 years; 910 with 9–11 years; and 438 with 12 years. Although low achievers should be expected to have lower levels of incomes than higher achievers, the rate at which income *increases* with respect to *additional* schooling or learning may or may not be smaller for the low achievers; it could be even greater.

The parameter estimates in Model I (Table 1) provide the type of information from which many have concluded—but too readily—that schooling pays off finan-

[10] The zero-order correlation coefficients are shown in Appendix Table A, with the means and standard deviations in Appendix Table B.

cially for low achievers. Level of educational attainment (*LEA*), X_1, is indeed highly significant, with a coefficient suggesting that each additional year of schooling (around the mean of 8.9) contributes an additional $62 per year to earnings for this low achievement group. Age, too, is significant, with each year (around the mean of 21.9)—presumably reflecting added experience and maturity—being associated with an additional $191 per year in earnings. Especially striking is the $608 average difference in earnings between whites and nonwhites. Equally notable, though, is the low R^2 shown in column 10, namely, 0.107 indicating that factors other than differences in *LEA*, age, and race, must explain the remaining 89 percent of the observed variance in annual earnings.

Model II introduces, as X_2, the learning variable, *AFQT* percentile score. Recall that we hypothesized that learning, not formal schooling per se influences earnings. As expected, the results show that *AFQT* is highly significant.[11] It apparently takes over about half of the effect previously assigned to schooling, X_1, for we now find that years of schooling falls in significance and is reduced in size from $61 to $30. An additional year spent in school while learning little, does appear to increase earnings, but only by $30 per year; but if measures are taken so that the amount learned is increased, earnings will, it appears, rise by $27 per year for each additional percentile point on the *AFQT*. In the light of this finding, one may question the wisdom of campaigns to keep teen-agers in school, which do not change their attitudes and motivation, do not improve the quality of the instruction offered, and do not affect the amount of learning that takes place.[12]

In Model III we add a variable indicating whether or not the respondent received any vocational or apprenticeship training. With the addition of the variable X_3, the coefficient of the schooling variable, X_1, becomes insignificant. Training, however, appears to be a large and very important factor, adding $326 per year to earnings for those who have received training. Note, too, that whereas most of the vocationally-valuable content of ordinary schooling is captured by the *AFQT* (previously seen in the comparison of Models I and II), *AFQT* picks up only a small portion of the effects of the training variable in Model III. A comparison of Model III with Model IIIa shows to what degree the financial return from training, like that from schooling, depends heavily upon the kinds of learning that are measured by the *AFQT*. Model IIIa includes the job training variable but not *AFQT*. The results indicate that exclusion of the learning variable, *AFQT*, has a small relative effect on the coefficient of the training variable, but a large relative effect on the coefficient of *LEA*, X_1. We conclude, therefore, that the vocational value of training seems to stem from types of learning not measured by *AFQT*.

Finally, in Model IV we include all of the variables that prior theorizing led us to believe are relevant, and for which we have data. These include variables X_6–X_9, reflecting current and past family circumstances and current economic opportu-

[11] The use of the *AFQT* percentile score rather than the raw score was not expected to affect the results substantially since for the range of scores in our sample, the percentile scores are approximately a linear transformation of the raw scores. Presented here is a counterpart of Model II using the *AFQT* raw score as X_2. Similar results were obtained using the raw score in Models III and IV.

Constant	Education	*AFQT* Raw Score	Age	Color	R^2
196.1	27.3**	13.6*	203.1*	539.7*	-
(134.3)	(13.2)	(2.2)	(15.2)	(59.0)	0.121

[12] See Weisbrod 1965.

nities. The estimates of Model IV, also shown in Table 1, disclose that $AFQT$, X_2, training, X_3, age, X_4, and color, X_5 retain their significance. But schooling remains insignificant, and the absolute size of the coefficient, $20 per year in additional earnings per year of schooling, would appear unimportant even if it were statistically significant. In addition, each added year of schooling contributes $24 in annual income through the apparent effect of education on $AFQT$ and the subsequent effect of $AFQT$ on earnings. Thus, the total effect of a year of schooling is estimated at $44. The $24 figure reflects our estimate that each additional year of schooling for this group contributes approximately one point to $AFQT$, and each additional $AFQT$ point is worth $24 per year.[13]

Of particular interest here are the effects of the variables reflecting family circumstances and economic environment. The coefficient of the current marital status variable, X_6, indicates as predicted, that being married is associated with the man's having additional earnings of $459 per year (significant at the .01 level). Coming from a home with divorced parents, X_7 or from a large family, X_8, is associated with somewhat lower earnings as expected, but the coefficients are not significant.[14] The region variable, X_9, has a large (and significant) coefficient, which shows that there is considerable financial advantage associated with living outside the South.

Not surprisingly, none of the models appears to explain a substantial portion of the cross-section earnings differentials; this is indicated by the relatively low values of R^2. Several reasons can be given for this. First, the men in the sample are too young to have settled into their lifetime work patterns so that the full impact of many of the variables does not yet appear. Second, earnings information is available for only a single year while the appropriate earnings variable is probably an average value for a period long enough to eliminate transitory factors such as illness and job changes. And finally, the youthfulness of the sample implies a lack of seniority which makes them subject more to seasonal and cyclical unemployment.[15]

Whatever the reasons for the low R^2 value, the fact remains that formal schooling, LEA, apart from its influence on learning, $AFQT$ is weaker than commonly supposed in explaining earnings differentials. We find that the overwhelming bulk of earnings differentials, at least those existing in a single year and still early in the careers of these low achievers, cannot be attributed to schooling, even when the effect of schooling on $AFQT$ is taken into account.

Our data and findings can cast some light on the hypothesis that there exists a "sheepskin effect" such that the mere possession, *ceteris paribus*, of an educational degree, for example a high school diploma, brings additional financial returns.

[13] This result was obtained by regressing $AFQT$ on the independent variables, including LEA, in Model IV. Specifically, the value of schooling was determined by regressing $AFQT$ on the independent variables in Model IV to determine the earnings value of an additional year of schooling.

[14] The lack of significance of the family-size variable, X_8, in our multiple-regression analysis is in notable contrast to the findings from a simple correlation analysis. Thus, the Task Force Report points out that 47 percent of rejectees came from large families, with five or more children, (p. A23).

[15] An indication of the importance of these latter two sets of factors is the increase in R^2 to 0.360 which was obtained by including a weeks-worked variable with the other explanatory variables of Model IV. Although this variable did serve as a proxy for these transitory factors as well as for motivation, its inclusion in our final model did not seem justified, since work-time might also reflect employers' discrimination among workers of differing educational and ability levels. Inclusion of the weeks-worked variable had no substantial effect on the coefficients or significance levels of the other independent variables.

A test of this hypothesis involves determining whether men who graduate from high school have higher earnings than men who do not, holding constant at least the level of learning (*AFQT* score), and age. The desirability of making such a test stems from the hypothesis that there is information value to employers in the knowledge that an individual graduated from high school. This information value arises from the consistency of effort required by a student to complete assignments, follow instructions, and get along reasonably with superiors and other students for four years. In contrast, an employer knows little about the ability of a dropout to meet such demands.

The test we propose involves determining whether the mean earnings of high school graduates (those persons having 12 years of schooling) are significantly greater than those of high school dropouts (having 9–11 years of schooling) of the same age with the same *AFQT* scores. Our findings are that the mean earnings for the two *LEA* groups were not significantly different. Thus, this test does not support the sheepskin hypothesis that graduation per se increases earnings, at least not for our low-achievement groups.[16]

In discussing the impact of training, we have been constrained by the serious lack of information on the amount or kind of training that respondents possessed. But this does not prevent us from making some crude estimates of lifetime payoffs associated with training relative to schooling. On the assumptions that the observed absolute financial payoffs from training will persist over the remainder of the working lifetime, and that the market rate of return on investment is, alternatively, 5 percent or 10 percent, we can capitalize the additional earnings, $291 per year, over the expected working lifetime to estimate the value of training. We arrived at figures of about $5,300 (at 5 percent) and $2,900 (at 10 percent). But now consider the payoff to added schooling. Capitalizing the additional earnings stream ($44 per year) associated with one additional year of schooling, resulting both directly through schooling and indirectly through the effect of schooling on *AFQT* and, then, on earnings, yields a present value of from $800 (5 percent) to $440 (10 percent). When we compare this to the individual's cost of acquiring one more year of schooling, anywhere from $1,500 to $3,000 per year of income foregone, the estimated payoff to more schooling is low for our sample group of low achievers. Thus, in terms of the net lifetime payoffs, training appears to be considerably more attractive than more schooling.

III. *Summary*

The results presented here demonstrate the misconception of postulating highly significant simple relationships between schooling and earnings, especially for low achievers. Though incomplete, the results indicate some directions for future inquiries. The naive approach which regards schooling as the only important determinant of earnings within age-sex-race groups is very incomplete, as is indicated by the low value of R^2—less than 11 percent—in the Model I in which schooling level, age, and color are used to explain variations in earnings. Such a result leads

[16] The fact that the hypothesis is not supported may come as no great surprise considering the special nature of the sample. Since all its members scored within a limited range of the *AFQT* exam—below the 30th percentile—it may be that those who were able to achieve this score with less schooling had more ability or a better background. These factors could have overridden the sheepskin effect, if it exists. But this is purely speculation, and thus additional testing for the presence of the sheepskin effect is needed.

us to hypothesize that earnings are a function of many other factors, and to an attempt to specify and quantify those factors. It also leads us to question whether schooling is itself one of these factors, or whether its apparent significance is due to its correlation with some other more fundamental variables.

The significance of years of schooling per se diminishes with the introduction of a measure of learning, such as the score on the *AFQT*. In short, it is apparently true that what one learns, in or out of school, influences earnings more than does the mere fact of spending time in school. A more fundamental approach, which we could not take because of data limitations, would examine the factors affecting learning, i.e., ability, motivation, home environment and quality of schooling, to determine their individual and collective roles.

The distinction between the importance of schooling and of learning raises a question about the policy of encouraging low-achiever students to remain in school. They are unlikely to benefit financially unless an attempt is made to insure that they *learn* in school rather than merely *attend* school, and that they are not deprived of other valuable opportunities, such as training programs, to enhance their earning power. Our results show rather dramatically the comparative payoffs to schooling and training, although we have no information concerning the nature or duration of the training programs. The returns to training are large enough, however, to suggest the wisdom of expanding training facilities rather than simply urging school attendance.

The most complete model we presented explained only 15 percent of the variance in annual earnings among the men in the sample. No doubt this is partly a result of limiting the number of variables; there were other variables that we believed to be significant determinants of earnings but which we were unable to estimate with the available data.

Appendix

Table A—Correlation Coefficients

Earnings	Education	*AFQT*	Training	Age	Color	Marital Status	Divorce of Parents	Family Size	Non-South	
1.000	.048	.132*	109*	.252*	.202*	.157*	−.042**	−.075*	.204*	Earnings
	1.000	.376*	.103*	−.048*	−.194*	−.014	.003	−.141*	.013	Education
		1.000	.086*	−.115*	.105*	.031	.030	−.152*	.088*	*AFQT*
			1.000	.044**	.076*	.022	−.037	−.065*	.069*	Training
				1.000	.067*	.173*	−.066*	−.013	.007	Age
					1.000	.018	−.130*	−.140*	.327*	Color
						1.000	−.019	.000	.007	Marital Status
							1.000	−.112*	.054	Divorce of Parents
								1.000	−.125	Family Size
									1.000	Non-South

* Indicates the coefficient is significantly different from 0 at the 0.01 level.
** Indicates the coefficient is significantly different from 0 at the 0.05 level.

TABLE B—MEANS AND STANDARD DEVIATIONS OF VARIABLES

Variable	Mean	Standard Deviation
Earnings, Y	\$1776.90	1480.30
Education, X_1	8.94	2.40
AFQT, X_2	9.11	6.91
Training, X_3	0.11	0.32
Age, X_4	21.90	1.88
Color, X_5	0.47	0.50
Marital Status, X_6	0.14	0.35
Divorce of Parents, X_7	0.31	0.46
Family Size, X_8	0.47	0.50
Non-South, X_9	0.40	0.49

REFERENCES

G. S. Becker, *Human Capital*, New York 1964, pp. 79–90.

M. E. Borus, "A Benefit-Cost Analysis of the Economic Effectiveness of Retraining the Unemployed," *Yale Econ. Essays*, fall 1964, *4*, 371–430.

D. S. Bridgman, "Success in College and Business," *Personnel J.*, June 1930, 4, 1–19.

E. G. Denison, "Measuring the Contribution of Education (and the Residual) to Economic Growth," in *The Residual Factor and Economic Growth*, Paris 1964, pp. 86–100.

V. R. Fuchs, *Differentials in Hourly Earnings by Region and City Size, 1959*, Nat. Bur. Econ. Res., Occas. Pap. 101, 1967.

S. J. Hunt, "Income Determinants for College Graduates and the Return to Educational Investment," *Yale Econ. Essays*, fall 1963, *3*, 305–57.

B. D. Karpinos, "Results of Examination of Youths for Military Service, 1965," *Supplement to Health of the Army*, July 1966, p. 6.

———, "The Mental Test Qualification of American Youths for Military Service and its Relationship to Educational Attainment," *Proc. Amer. Statist. Assoc.*, 1966, pp. 92–111.

J. Mincer, "On-The-Job Training: Costs, Returns, and Some Implications," *J. Polit. Econ.*, Oct. 1962, *70*, pp. 50–79.

D. A. Page, "Retraining Under the Manpower Development Act: A Cost-Benefit Analysis," *Public Policy*, 1964, *53*, 257–67. The Brookings Institution No. 86.

T. Ribich, *Education and Poverty*, The Brookings Institution, Washington 1968.

G. G. Somers and E. W. Stromsdorfer, "A Benefit-Cost Analysis of Manpower Retraining," *Ind. Relat. Res. Assoc. Proc.*, Dec. 1964, pp. 172–85.

B. A. Weisbrod, "Preventing High School Dropouts," in R. Dorfman, ed., *Measuring Benefits of Government*, The Brookings Institution, Washington, 1965, 117–49.

——— and P. Karpoff, "Monetary Returns to College Education, Student Ability, and College Quality," *Rev. Econ. Statist.*, Nov. 1968, *50*, 491–97.

D. Wolfle and V. G. Smith, "The Occupational Value of Education for Superior High School Graduates," *J. Higher Education*, April 1956, *27*, 201–14.

President's Task Force on Manpower Conservation, *One-Third of a Nation*, Washington, Jan. 1, 1964.

U.S. Bureau of the Census, *1960 Census of Population, Subject Reports, Occupation by Earnings and Education*, Series PC(2)–7B.

A Simple Algebraic Logrolling Model

By Gordon Tullock*

In the rapidly developing literature in which essentially economic tools are applied to political problems, there have been two major models of voting performance. One of the models, by all odds the most widely used, is essentially spatial. In it, individuals are assumed to have a preference mountain and to prefer the points which are closer to their optimum to points which are farther away. This model, which started as a very simple one-dimensional continuum in the work of Harold Hotelling, Duncan Black (1948), and Anthony Downs, has developed into a more complex, many dimensional model in the later work of Black (1958), Otto Davis and M. J. Hinich (1966, 1967, 1968), and Tullock. The many dimensional version of this model must be represented, of course, by some variant on the Cartesian algebra since it is not easy to represent graphically more than two dimensions on a piece of paper. In general, these models have been used mainly to demonstrate that in a two-party system, the two parties will normally have platforms that are very similar and that these will represent median preferences. The other model deals with the phenomenon of logrolling and has normally been represented by other tools, see James Buchanan and Tullock. The interrelation between these two models has been discussed in general by Davis and Hinich (1968, p. 68) and Tullock (pp. 57–61), but no very rigorous joint model exists. It is the purpose of this article to demonstrate that the two approaches are not inconsistent by presenting a spatial model which will also cover logrolling.

I. *Three-Person Model*

Suppose that an individual must choose government policies with respect to three different issues which we shall designate A, B, and C, and that each of these issues represents a continuum, such as the appropriation for the army or the appropriation for the welfare program. This situation can be represented by a three-dimensional issue space with each of the issues representing one dimension and an individual having some point which is for him optimal, let us say [10, 10, 10]. Presumably his level of satisfaction will fall off as the actual social choice moves away from his optimum. If we assume that this fall-off is uniform in every direction, we may express his loss from not achieving his optimum by the equation:

$$(1)\quad L_A^2 = [A - 10]^2 + [B - 10]^2 + [C - 10]^2$$

If we had a number of people with varying optima in the issue space, we would be able to deduce from the resulting set of equations similar to equation (1) how they would vote on each proposition that was put before them.[1] As has been demonstrated by the spatial models so far published, except with very special distributions of optima, the outcome under simple majority voting would be some point which is approximately at the median of the entire distribution. This conclusion is readily generalizable up to any number of issues, since the Cartesian system can be applied to an issue space of any number of dimensions.

The use of perfectly spherical indiffer-

* The author is professor of economics and public choice at Virginia Polytechnic Institute.

[1] See Colin Campbell and Tullock for an application of this process.

ence surfaces in this model does not appear to restrict its utility very much. In the real world, we would not anticipate such perfection, but the deviations from it would be essentially random and the law of large numbers should lead, where there are many voters, to much the same outcome as if we used our spheres. For a demonstration, see Kenneth Arrow. Systematic deviations from the spherical model, together with appropriately structured locations of the individual optima, could lead to voting cycles, and the conclusion that the median preference dominates would be undermined. It is the purpose of this article to consider an important case in which we would anticipate that the individual indifference curves would systematically vary from the spherical in a particular way, and in which we would anticipate that individual preferences would have a structure such that the combination of these two effects leads to quite different results than have customarily been dealt with by the spatial models.

If we consider those situations in the real world in which we observe logrolling and compare them with those situations where logrolling appears to be relatively unimportant, we observe immediate differences in the structure of the individual preferences. In logrolling, we observe a number of people who are highly interested in one particular project, let us say, the dredging of the James so that Richmond becomes a deep water port, and only mildly interested in other projects which, generally speaking, they oppose. The rivers and harbors area is the *locus classicus* of logrolling, but similar phenomenon will be found throughout a very large part of modern governments.

The indifference curves of the individuals engaging in logrolling are somewhat similar to those shown on Figure 1. Mr. A wants his harbor dredged at the expense of the general taxpayer and feels quite

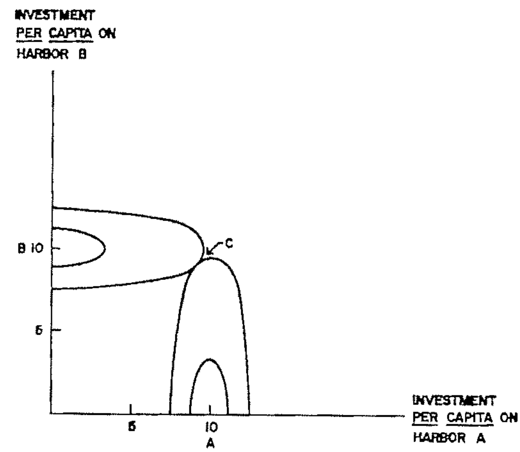


FIGURE 1

strongly about it, but he would rather not pay for dredging Mr. B's harbor. Since he is only one of many taxpayers, however, his feeling about the dredging of Mr. B's harbor is much feebler than his feeling about his own. Mr. B's feelings are the converse. If we assume that all of the citizens of the town in which Mr. A lives feel much the same as Mr. A[2] and the other citizens of the town in which Mr. B lives feel much the same as Mr. B, then logrolling becomes rational. Point *C* is better than the origin for both A and B. It is not, however, possible to represent a many dimensional logrolling process in two dimensions and, if we consider such a piece of legislation as the rivers and harbors bill, it is clear that several thousand dimensions would be necessary.

We can begin with a simple three-dimensional model using the ordinary Cartesian algebra. This model for logrolling will differ from the usual spatial model only in that the individuals are assumed to have intense preferences on certain subjects. If we assume a three-person society where

[2] This does not mean that their optima or intensities are the same as Mr. A, but simply that all of their indifference curves would have the same general shape as Mr. A's.

harbor dredging is paid for by equal per capita taxes, and that there are three harbor dredging operations contemplated (A, B, and C), then Mr. A's preferences can be represented by the first equation in Table 1. Mr. A is assumed to have his personal optimum at the point [10, 0, 0]. L_A is the "loss" he suffers if the government chooses some other point. As in the real world, he is much more interested in the dredging of his own harbor than in preventing the dredging of the other two harbors, although he doesn't like paying taxes to benefit other people. Note that Mr. A's optima includes his own per capita share in terms of tax payments for his own harbor. For simplicity we will continue to assume throughout that all expenditures on logrolled projects are paid for by a tax which is evenly divided among the taxpayers and the per capita cost is our metric on each issue dimension.

II. *Logrolling Results*

Taken on the two dimensions represented by axes A and B and holding C equal to zero, the indifference curves generated by the equations of A and B in the positive quadrant will approximate those shown in Figure 1. In three dimensions, A's indifference surfaces in the positive part of the issue space would form a quarter of a disk with its center at the point [10, 0, 0]. The other three individuals in our current simple model would have similar disks attached to the other three axes. If the voting rule is simple majority voting, and each individual votes for his preference on each of the three harbor proposals, then there will be two votes against each proposal and all will fail. L_A would equal 22.4. The individuals, however, should notice the possibility for gains from trade. If two of them could get together and vote for each other's harbor dredging project, then they can make a quite considerable gain. Bargaining difficulties in this case are apt to be minimal since each party to the bargain has the alternative of turning to the third party and hence in essence they are operating in a market type situation.[3] Thus, if we assume that there is some agreement between Mr. A and Mr. B and that they choose an equal amount of harbor dredging in each of their harbors,[4] it's fairly easy to determine the point in the issue space which would result. It is [8 1/3, 8 1/3, 0]. L_A and L_B will be 9.1,[5] very much better than the situation without the agreement. L_C, on the other hand, is now 27.2, much worse than the situation before the agreement was made. The reason, of course, is simple. Mr. C's harbor is not being dredged and he is paying taxes to dredge the other two.

This may be taken as a very simple example of the type of bargain which occurs in logrolling. In practice, things are more complicated. There are basically two types

TABLE 1

$$L_A^2 = 5(A-10)^2 + B^2 + C^2$$
$$L_B^2 = A^2 + 5(B-10)^2 + C^2$$
$$L_C^2 = A^2 + B^2 + 5(C-10)^2$$

[3] A modern theory of bargaining can be said to have been initiated by J. von Neumann and O. Morgenstern. For a summary of developments since publication of *Theory of Games and Economic Behavior* and some interesting experimental results in a three-person situation, see William Riker (1967).

[4] The assumption of equal division is not strictly necessary for the general conclusions reached below. The equations would also be solvable for other assumptions as to the division of the spoils between the members of the logrolling bargain.

[5] The derivation of these results is simple but not obvious. Since A is equal to B, and C is equal to 0, the first equation in Table 1 reduces to: $L_A^2 = 6A^2 - 100A + 500$. Differentiating [the variable L_A^2 with respect to A] and setting the differential equal to 0 produces a value of A equal to 8.33 which minimizes both L_A^2 and L_A. Substitution of this value into the equations of Table 1 gives the numbers shown. Similar methods are used for all further computations in this article. The computations were carried out on a slide rule, which is rather unusual in these days of computers, and hence there is a possibility of error in the final decimal.

of logrolling. The first is explicit logrolling, most often observed in Congress although it does occur in other situations. It involves individuals who trade their votes on many individual issues to many others for votes on other issues. Under these circumstances, there is no reason why any particular person would be left out. Everyone may trade with anyone else and the result amounts to a peculiar market solution. Since in making the trades, each individual is only attempting to make up a majority coalition, the cost calculations are similar to those informing the agreement we just discussed. Nevertheless, the fact that everyone may get their project means the outcome is different.[6] With just three voters this result would not occur, but with more voters the outcome of this process might well be that all of the harbors would be dredged at a level equivalent to 8 1/3.[7] The individual is in marginal adjustment on those logrolling deals in which he has participated and loses on those in which he has not. L_A, assuming that all three harbors are dredged to the level of expenditure of 8 1/3, is 12.3. This is worse than Mr. A obtains from his simple agreement with Mr. B, but it is certainly much better than he would obtain if no agreements were made at all.

The other type of logrolling is called implicit logrolling and involves political parties or candidates who present "platforms." These platforms, in essence, are complex mixes of different measures. A proposal to dredge two of three harbors would be an example of such a platform. Assuming that this type of logrolling is adopted then the individual *ex ante* has two chances out of three of being a member of the coalition and having his harbor dredged, and one chance in three of having to pay taxes for the dredging of two other harbors. Discounting this out in a simple manner, L_A *ex ante* would be 15.1. Once again, it is much better than a no-logrolling solution. Unfortunately, this type of solution is not mathematically stable, but we will defer discussion of the matter, closely related to Arrow's general impossibility theorem, until the latter part of this article.

So far, we have said nothing about Pareto optimality. If we require unanimity, clearly the bargaining costs would be high, but the economists would normally anticipate that the ultimate outcome, if we disregard the bargaining costs, would be better than the outcomes obtained by partial agreement. The Pareto optimal area is, of course, quite a complex surface running across the three-dimensional space. We can, however, fairly easily compute the value of one particular point on that surface. With our highly symmetric model, side payments would lead to a decision to dredge all three harbors equally at 7 1/7 each and the loss function of that point to Mr. A would be 11.8. *Ex ante* the side payments would cancel out and this is better than any of the other possibilities we have discussed. Needless to say, for an individual who can feel sure that he will be one of a pair of voters who have only their harbors dredged, that outcome would be better than the Pareto optimal outcome.

III. *Five Voters*

The method of calculation we have been describing can readily be applied to any

[6] See James Coleman for a discussion of this point. His article led to comments by R. E. Park and Dennis C. Mueller which, together with a reply by Coleman, are printed in the Dec. 1967 issue of this *Review*, pp. 1300–16.

[7] The simplest way of understanding this problem is to assume that an individual purchases other peoples' votes with his own. There is no obvious reason, if there are more than three voters, why the votes purchased by Mr. A should be the same votes as those purchased by say, Mr. C, although Mr. A's collection of purchases includes Mr. C's vote and Mr. C's collection of purchases includes Mr. A's vote. Mr. A could, for example, make up his majority in a 5-man voting system out of *A*, *B*, and *C*, and Mr. C make up his out of *C*, *D*, and *E*. Mr. E, similarly, might have a majority which consists of C and B as well as himself.

number of dimensions. For example, assume that there are five harbors and five voters (groups and voters) whose loss functions are as shown in Table 2. Once again, if all of the projects for dredging harbors are put up individually and all the individuals vote on them strictly in accordance with their preference on each issue, in each case there will be four votes against and one in favor. The resulting outcome will be the origin of the five-dimensional Cartesian axis system and L_A will be again 22.4. If we assume that three groups of voters, those on harbors A, B, and C, get together to form a majority, they would agree to vote for [7 1/7, 7 1/7, 7 1/7, 0, 0]. This gives L_A equal to 12.0, much better than would be obtained without bargaining. Messrs. D and E, not members of the winning coalition in this case, however, find that the payoff of 25.6 is worse than would have been obtained had logrolling not existed.

Using our group of five, however, we can consider a variety of voting rules. Table 3 shows on the left the minimum size of the coalition which is required by various voting rules. The outcome in terms of the amount of dredging in each harbor is shown in the second vertical column, the third column shows the payoff to a member of the winning coalition, and the fourth, the payoff to a man who is left out. In the final column, we show the *ex ante* value of the arrangement for some person who does not know whether he will be a winner or a loser but who assumes his probability of being in the winning coalition is proportional to the number of people required. For comparison purposes, we have put the no-logrolling outcome at the bottom of the table.

TABLE 2

$$L_A^2 = 5(A-10)^2+B^2+C^2+D^2+E^2$$
$$L_B^2 = A^2+5(B-10)^2+C^2+D^2+E^2$$
$$L_C^2 = A^2+B^2+5(C-10)^2+D^2+E^2$$
$$L_D^2 = A^2+B^2+C^2+5(D-10)^2+E^2$$
$$L_E^2 = A^2+B^2+C^2+D^2+5(E-10)^2$$

The reader way be surprised at the existence of a voting rule permitting a coalition of two to obtain the dredging of their harbors. It is not, however, an unrealistic situation. Most modern democracies use a representative assembly. Under these circumstances, a majority of the voters in a majority of the constituencies may be able to control the outcome. Thus, less than a majority of the voters is necessary. Our two-voter coalition is an example.[8]

It will be noted that the numerical outcomes we have obtained from our simple calculation procedure are in exactly the form which would have been predicted

[8] The proportional representation system used so much on the continent of Europe generally speaking makes it impossible for less than the majority of the voters to have the influence shown. In Anglo-Saxon countries, however, the possibility does exist for the minority of voters obtaining the type of profits shown here.

TABLE 3

Minimum coalition size	Platform	Payoff to member of winning coalition	Payoff to nonmember	*Ex ante* Payoff
2	8 1/3,8 1/3,0,0,0	9.1	27.2	19.9
3	7 1/7,7 1/7, 7 1/7,0,0,	12.0	25.6	17.4
4	6 1/4,6 1/4, 6 1/4, 6 1/4, 0	13.7	25.6	16.1
5	5 5/9,5 5/9, 5 5/9, 5 5/9, 5 5/9	15.0	—	15.0
No logrolling	0, 0, 0, 0, 0			22.2

from the nonnumerical discussion in Buchanan and Tullock. The costs of coalition formation, of course, must be offset against the numbers in Table 3 to find the optimal voting rule. The point of this model has not been to advance the line of reasoning started in Buchanan and Tullock. Instead it provides a basis for future research by demonstrating that it is possible to obtain their conclusions through a model which differs from the widely used spatial models only by a minor change in parameters.

The outstanding characteristic of the type of issue that normally involves logrolling as opposed to the type of issue that normally does not, is simply that there are groups of voters who feel much more strongly about one particular issue than about others, and that these different groups of voters are arranged roughly in the symmetrical way that we have shown. Needless to say, the perfect symmetry which I have given the model is an aid to calculation, not an effort to describe the real world.

IV. *Other Models*

In order to move from the model we have here to the type of model that was used in Davis and Hinich (1966, 1967, 1968), and Tullock, we may begin by assuming that the individuals favor all of the goods provided to some extent. Suppose for example, that individual A's preferences for the dredging of harbor B is not simply an aversion to taxation for this purpose but that he actually does think it would be nice to have it dredged. Under these circumstances, the center of the disk which now describes his loss function would be moved away from the A-axis a short distance and corresponding computations would indicate that there would be somewhat more dredging of harbor B. This could also lead to the ellipse being shorter and fatter. In the limit, if we continue such operations, we would end with a circle with its center somewhere near the middle of the issue space.

However, we do not have to change our loss functions from disks to spheres in order to obtain approximately the results obtained by the analysis which shows the central policy is dominant. All that is necessary is to relax our extremely strict restrictions upon the shape of individual preferences. We have grouped the individuals in clusters along the axes very strongly favoring certain projects which benefit them and being opposed, mainly because of the tax cost, to individual projects of the same nature in other areas. This is, indeed, a very tight restriction. Unfortunately, it would appear that it is very commonly met in the real world. If we assume that this type of clustering does not occur, then we are back in the world of Davis and Hinich, and Tullock. Thus, we have obtained logrolling essentially out of the spatial model simply by assuming that there are people with the type of preference that we observe in logrolling situations.

So far, however, we have assumed that our function is stable. In actual fact, what we have referred to as explicit logrolling is indeed a stable situation, but, what we have called implicit is not. For example, if we return to the set of equations in Table 1, the platform [10, 10, 0] can be beaten by the platform [10, 0, 0] which can be beaten by [0, 0, 0] which in turn can be beaten by [10, 10, 0]. Further, [10, 10, 0], [0, 10, 10], [10, 0, 10] are all possible winning outcomes. In my *Towards a Mathematics of Politics*, I argued that the instability (implied by the Arrow theorem) in respect to voting was of little real importance. My demonstration, however, depended on the assumption that the number of voters was very much in excess of the number of issue dimensions. When the voters are clustered well out on each of the issue dimensions as they are in our logrolling

model, the proof that I offer ceases to be relevant. In essence, each cluster of voters acts as one voter and the number of such clusters is the same as the number of issue dimensions.

V. *Applications*

In the real world, voting would appear to cover many issues in which the preferences of the individual voters do not have the high degree of structure required for logrolling issues. The classical solution for such a problem for a party wishing to maximize votes would be to seek the middle position on the nonlogrolling issues, and on the logrolling issues, attempt to seize a position which is superior to whatever his opponent has offered. This would lead to sharp changes of policy and great differences between the two parties. We do not observe either of these things in the real world.

It should be noted that a good deal of the logrolling actually done in Congress is on an explicit basis rather than by the parties on an implicit basis. Both the Republican and Democratic congressional candidates from Richmond will be in favor of dredging the harbor. Both will also be against (although not very strongly) dredging other harbors. When they get into the House, the explicit bargaining scheme which is stable will explain their behavior. Unfortunately, there are many types of logrolling which take place at the platform level and hence, the instability problem still remains.

Why do we not see this kind of change in the real world? One possible solution is simply that without the high degree of symmetry which I have imposed upon my model there may be genuinely superior coalitions. Riker (1962) discussed one particular set of conditions under which certain coalitions are "better" than others. There may be many other similar situations.

This solution, however, obviously has its drawbacks and I think we can construct another solution which is both simpler and closer to the real world. In his recent article, Arrow pointed out that "Since the effect of any individual vote is so very small, it does not pay a voter to acquire information unless his stake in the issue is enormously greater than the cost of the information." These theoretical considerations which indicate that people should not bother to become informed about politics can be matched with empirical data which seems to indicate that they do not, in fact, know much about politics. If we assume that individuals will only make an effort to find out about policies when the effect on them is greater than a certain amount, then the individual would normally know, at least, something about what we might call public interest issues, such as police and national defense and also something about those logrolling issues which particularly concern him, i.e., those upon which his feelings are intense. However, he would not know anything about those logrolling issues which did not greatly affect him.

All the inhabitants of Richmond would know about the James River dredging project but few of them would know about the dredging of the river to Tulsa, Oklahoma. Under these circumstances, a political party making up its platform would assume that different voters have somewhat different information positions. In the extreme, the voter in Richmond would respond to a political world in which he saw general issues and the dredging of the James. In this area of his information there will be no possibility of strict logrolling and hence both political parties would choose approximately the center of this issue space. With our assumptions of voter ignorance of other issues, this would involve dredging the James to the level of 10. With many voters in Richmond, the point chosen would be the median of *their* pref-

erences. The party also assumes similar positions with respect to other electorates which have different fields of knowledge and preferences. The outcome would involve logrolling in a sense that the individual groups would be given special treatment but would depend upon the ignorance of the voter with respect to the logrolling "payments" to other parties. Whether voters actually *are* this ignorant is something which can be questioned. Certainly they are opposed to taxes in general, and are aware of the fact that other people's projects, in one way or another, contribute to the tax load. The empirical investigations which do show appalling voter ignorance have never been addressed to this specific problem. Further empirical research would appear to be called for.

REFERENCES

K. Arrow, "Tullock and an Existence Theorem," *Publ. Choice*, spring 1969, *6*, 105–11.

D. Black, "On the Rationale of Group Decision Making," *J. Polit. Econ.*, Feb. 1948, *56*, 23–34.

———, *A Theory of Committees and Elections*, Cambridge 1958.

J. M. Buchanan and G. Tullock, *The Calculus of Consent*, Ann Arbor 1962.

C. Campbell and G. Tullock, "Computer Simulation of a Small Voting System," *Econ. J.*, forthcoming.

J. S. Coleman, "The Possibility of a Social Welfare Function," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1105–22.

O. A. Davis and M. J. Hinich, "A Mathematical Model of Policy Formation in a Democratic Society," in J. Bernd, ed., *Mathematical Applications in Political Science II*, Dallas 1966.

——— and ———, "Some Results Related to a Mathematical Model of Policy Formation in a Democratic Society," in J. Bernd, ed., *Mathematical Applications in Political Science III*, Charlottesville 1967.

——— and ———, On the Power and Importance of the Mean Preference in a Mathematical Model of Democratic Choice," *Publ. Choice*, fall 1968, *5*, 59–72.

A. Downs, *An Economic Theory of Democracy*, New York 1957.

H. Hotelling, "Stability in Competition," *Econ. J.*, Mar. 1929, *39*, 41–57.

W. Riker, *The Theory of Political Coalitions*, New Haven 1962.

———, "Bargaining in a Three Person Game," *Amer. Polit. Sci. Rev.*, Sept. 1967, *61*, 642–56.

G. Tullock, *Toward A Mathematics of Politics*, Ann Arbor 1966.

J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton 1944.

COMMUNICATIONS

The Revenue Maximization Oligopoly Model: Comment

By Milton Z. Kafoglis and Robert C. Bushnell*

A recent attempt by Robert Haveman and Gilbert DeBartolo in this *Review* to generalize the constrained revenue maximization equilibrium has demonstrated that an earlier formulation by Robert Sandmeyer is a special case. However, the Haveman-DeBartolo analysis does not correctly relate mathematical and geometrical models, leaving (in the case of the geometrical analysis) the erroneous impression that the revenue maximizing equilibrium is characterized by the equality of marginal cost and marginal revenue. Moreover, the explication of the mathematical model is, in several respects, incomplete. Therefore in this comment we bring attention to the error and further generalize this troublesome equilibrium.

Following Haveman-DeBartolo p. 1356) we define:

$x \cdot D(x, a) = TR(x, a) =$ total revenue function when x units are sold after a dollars of advertising expense.

$\pi =$ minimum acceptable profits

$C(x) =$ total cost of production function

$A(a) =$ total cost of advertising function

We maximize $TR(x, a)$ subject to the minimum profit constraint $TR(x,a) - C(x) - A(a) \geq \pi$ and with the "general" Baumol (p. 60) assumptions, $\partial TR/\partial a > 0$, $dC/dx > 0$ and $x > 0$. Using a Lagrangean form and the Kuhn-Tucker conditions we derive the following equilibrium conditions for revenue maximization, which are proved in the Appendix.

$$\text{(1a)} \quad \frac{dA}{da} \Big/ \frac{\partial TR}{\partial a} = \frac{dC}{dx} \Big/ \frac{\partial TR}{\partial x} = 1 + \frac{1}{\lambda}, \quad \text{when } a > 0$$

$$\text{(1b)} \quad \frac{dA}{da} \Big/ \frac{\partial TR}{\partial a} \geq \frac{dC}{dx} \Big/ \frac{\partial TR}{\partial x} = 1 + \frac{1}{\lambda}, \quad \text{when } a = 0.$$

When the firm is constrained,[1] $\lambda > 0$ and therefore, $1 + 1/\lambda > 1$. Thus, we have in the case of both (1a) and (1b)

$$\text{(2)} \qquad \frac{dC}{dx} > \frac{\partial TR}{\partial x}$$

Equation (1a) is a rearranged version of the statement given by Haveman and DeBartolo (p. 1356). However, they do not give equations (1b) and (2) or the $1+1/\lambda$ condition. We note: a) that marginal cost must exceed marginal revenue (equation 2); and b) that even though $\partial TR/\partial a > 0$, advertising need not be undertaken (equation 1b). These points require emphasis because the Haveman-DeBartolo graphical analysis leads to a (false) equilibrium in which marginal revenue and marginal cost are equated and in which surplus profits are devoted fully to advertising. Such an $MC = MR$ level of output cannot be optimal. A firm in such a situ-

* The authors are, respectively, professor of economics, University of Florida and assistant professor of economics, The Ohio State University. We gratefully acknowledge the assistance of William J. Baumol.

[1] Either Baumol's assumption that $\partial TR/\partial a > 0$, i.e. demand will always shift in response to advertising, or the assumption that $\partial TR/\partial x > 0$, i.e. demand is always elastic for any relevant output, will alone ensure that profits are constrained to the required minimum. Without at least one of these assumptions the possibility exists that the unbounded maximum $\partial TR/\partial a = 0 = \partial TR/\partial x$ could apply.

ation would find that some of the surplus profits expended on advertising would produce more revenue if expended instead on increased production and production expense. They would find such incremental transfers would continue to yield higher total revenue until either condition (1a) or (1b) was satisfied. If condition (1a) applies, profit surplus will be devoted to production as well as to advertising. If condition (1b) applies, no advertising at all will be undertaken. If the marginal contribution to revenue per additional dollar of production expense exceeds the marginal contribution to revenue per additional dollar of advertising expense for all combinations of output and advertising consistent with minimum profits, equation (1b) becomes relevant. In any case, price always will be lower and advertising expense smaller than when all profit surplus is allocated to advertising. In fact, since the relation $MC>MR$ may be rewritten as $MC>P\ (1-1/\eta)$ for price P and demand elasticity η, we cannot rule out the possibility that price could fall below marginal cost with $MC>P>MR$.

In conclusion, the model as formulated by Baumol incorporates the possibility of "excessive" production expense and lower price as well as the more widely recognized possibility of "excessive" advertising expense and higher price. Indeed the model is not inconsistent with strong price competition.

Appendix

Equilibrium Conditions For Revenue Maximization

The conditions are derived from the Lagrangian form L in the nonnegative variables x, a, and λ,

$$\text{(a)}\quad L(x, a, \lambda) = TR(x, a) + \lambda[TR(x, a) - \pi - C(x) - A(a)],$$

a concave objective function subject to a concave constraint. The conditions are

$$\text{(b)}\quad \frac{\partial L}{\partial x} = \frac{\partial TR}{\partial x} + \lambda\left[\frac{\partial TR}{\partial x} - \frac{dC}{dx}\right] \leq 0,$$

at maximum either (b) holds as an equality or else $x=0$,

$$\text{(c)}\quad \frac{\partial L}{\partial a} = \frac{\partial TR}{\partial a} + \lambda\left[\frac{\partial TR}{\partial a} - \frac{dA}{da}\right] \leq 0,$$

at maximum either (c) holds as equality or else $a=0$,

$$\text{(d)}\quad \frac{\partial L}{\partial \lambda} = [TR(x, a) - \pi - C(x) - A(a)] \geq 0,$$

at maximum either (d) holds as an equality or else $\lambda=0$.

Assuming $x>0$ but making no assumption about a, equations (b) and (c) may be rewritten as

$$\text{(e)}\quad (\lambda + 1)\frac{\partial TR}{\partial x} = \lambda\frac{dC}{dx}$$

$$\text{(f)}\quad (\lambda + 1)\frac{\partial TR}{\partial a} \leq \lambda\frac{dA}{da}$$

Assuming $\partial TR/\partial a>0$ implies the left side of (f) is strictly positive, hence λ cannot be zero without contradicting (f). Since λ is thus known to be strictly positive division by λ is permissible.

We therefore write

$$\text{(g)}\quad 1 + \frac{1}{\lambda} = \frac{dC}{dx} \bigg/ \frac{\partial TR}{\partial x}$$

$$\text{(h)}\quad 1 + \frac{1}{\lambda} \leq \frac{dA}{da} \bigg/ \frac{\partial TR}{\partial a}$$

Equations (1a) and (1b) follow from (g) and (h).

REFERENCES

W. J. Baumol, *Business Behavior, Value and Growth*, rev. ed., New York 1967.

R. Haveman and G. DeBartolo, "The Revenue Maximization Oligopoly Model: Comment," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1355–58.

R. L. Sandmeyer, "Baumol's Sales-Maximization Model: Comment," *Amer. Econ. Rev.*, Dec. 1964, *54*, 1073–81.

The Revenue Maximization Oligopoly Model: Comment

By C. J. Hawkins*

It is possible to show the solution to the general model of Sales Revenue Maximization in Oligopoly (W. J. Baumol, 1958, 1967) in such a way that we can show all possible levels of revenue, costs, advertising outlay, output, and profits all on one two-dimensional diagram.

First, however, it is easier to think of the Baumol model in three dimensions with revenue and costs on the vertical axis. These are plotted against output and advertising on two further axes.

Total revenue against output is assumed to be of conventional shape for any given level of advertising. Increases in advertising shift the total revenue curve upwards at first rapidly and then more slowly so that in effect what one has is as shown in Figure 1. There is a different total revenue curve (against output) for every level of advertising. The revenue surface can therefore be likened to the roof of a Nissen hut, which is small at one end (where advertising is zero) and large at the other end as advertising increases. Below the revenue surface would be the cost surface, like a floor below the roof of the Nissen hut. This cost surface would rise in one direction as output was increased and also in the other direction as advertising outlay was increased as well. A profit maximizer merely searches for the point where the gap between the floor and the roof (the cost and revenue function) is greatest. A revenue maximizer chooses a particular size of gap (his constraint level of profits) and then searches for

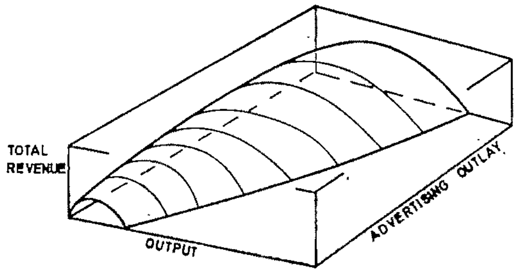


Figure 1

the position which gives him maximum total revenue for this level of profits.

Robert Haveman and Gilbert DeBartolo took a series of slices through this three-dimensional diagram (sliced through the advertising outlay axis) and superimposed them onto one two-dimensional diagram. Unfortunately this diagrammatic technique is of limited value since it can only be used to show a selected number of possible equilibria. As a result of this limitation, Haveman and DeBartolo made the error of suggesting that the solution to the Baumol model was the revenue maximizer's optimum from the possible positions shown in their diagram. But the diagram did not show that other total revenue curves must exist offering more revenue at the same level of profits.

I. *The Diagrammatic Solution: An Alternative Approach*

An alternative approach is as follows. Purely for simplicity, assume that average costs are constant and that there are no fixed costs; e.g., as in the long run. Then total costs with respect to output will be a straight line through the origin. Advertising outlay costs are measured by value. A locus of all points of equal outlay on various combinations of advertising and production costs must also therefore be a straight line. A series

* The author is lecturer in economics at the University of Southampton, England. A number of points that were originally intended to be made in this paper have been covered by Milton Kafoglis and Robert Bushnell in the current issue. Detailed proofs of these points therefore are not repeated in this paper but are available in the University of Southampton Discussion Papers in Economics and Econometrics (Number 6909).

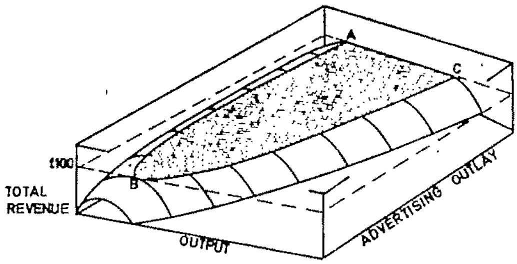


FIGURE 2

of these isocost curves can then be plotted for all possible levels of outlay. With constant average costs these isocost lines will be parallel and equidistant from each other (see Figure 3).

We can also plot the locus of all combinations of advertising and output that yield the same level of total revenue. This is most easily thought of as a slice through the roof of the Nissen hut (total revenue surface) at a given height, given level of total revenue. Such a slice through a total revenue surface is shown in Figure 2. The isorevenue curve for £100 of revenue is shown by the boundary of the shaded area (solid line ABC). Point B will always be at the peak of a total revenue curve whatever level of isorevenue is selected. (Otherwise it could not be a point but would be a series of points all yielding the same revenue, in this case, of £100. This could only happen if the peak of the relevant revenue curve were flat.) Note also that only the AB portion of the isorevenue curve is relevant since for any level of revenue on the section, BC, there is a similar level of revenue available on AB but at a lower total outlay.

We can now plot isorevenue and isocost curves for all levels of advertising and output that yield positive profits (Figure 3). The profit maximizer will go to point X with output Q_0 and an advertising outlay of £250. X is the point where the gap between costs (isocost line £500) and revenue (isorevenue curve £540) is greatest. He chooses a point of tangency between isocost and isorevenue curves because clearly he wants maximum revenue for his chosen outlay since profit represents the difference between the two.

The revenue maximizer also goes to a point of tangency since for his chosen outlay he also wants maximum revenue (in this case because he wants revenue for its own sake). At all tangency points the slopes of the cost and revenue functions show that

$$\frac{\partial TR}{\partial x} \Big/ \frac{\partial C}{\partial x} = \frac{\partial TR}{\partial a} \Big/ \frac{\partial A}{\partial a} \tag{1}$$

which means (as can be seen by cross-multiplying) that the *necessary* equilibrium condition for revenue maximization is met (Haveman and DeBartolo). This condition (1) is also necessary for profit maximization. Now although all tangency points meet this condition, only at point X is the condition met *and* profits maximized; i.e., only at X is the further necessary condition for profit maximization met that:

$$\frac{\partial TR}{\partial x} = \frac{\partial C}{\partial x} \quad \text{and} \quad \frac{\partial TR}{\partial a} = \frac{\partial A}{\partial a} \tag{2}$$

By contrast, a revenue maximizer with a constraint level of profits of £20 could go either to point W or point Y. Both satisfy the necessary condition (1), and both yield profits of £20. The revenue maximizer will, however, choose point Y since this offers the higher of the two possible levels of revenue and also meets the condition that:

$$\frac{\partial TR}{\partial x} < \frac{\partial C}{\partial x} \quad \text{and} \quad \frac{\partial TR}{\partial a} < \frac{\partial A}{\partial a} \tag{3}$$

which is necessary if the revenue maximizer is to increase revenue while reducing profits

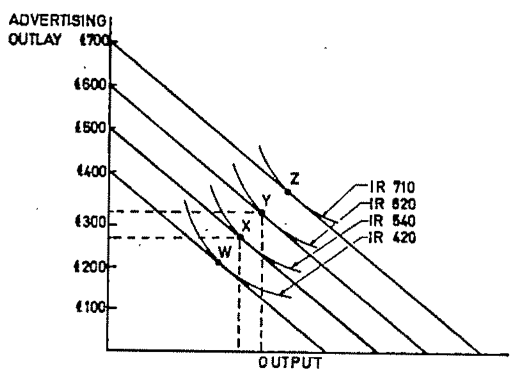


FIGURE 3

from the maximum level (where these terms would be equal) to his chosen constraint level. Point W does not meet this condition (3). At point W total outlay on output and advertising together is less than it would be at the maximum profit position: we know therefore that expansion from point W would raise revenue more than it would raise costs; i.e., it must therefore be true that at W:

$$(4) \qquad \frac{\partial TR}{\partial x} > \frac{\partial C}{\partial x} \quad \text{and} \quad \frac{\partial TR}{\partial a} > \frac{\partial A}{\partial a}$$

Point W cannot therefore meet the necessary condition (3). Point Y is therefore the equilibrium for the revenue maximizing firm. In this example, it will advertise more and produce a larger output than the profit maximizer.

With the aid of this diagram some of the controversial issues about revenue maximization can be solved to show that:

(a) The Sandmeyer equilibrium is correct for the special case of constant production costs.
(b) Baumol's claim that revenue maximizers in general produce a larger output than profit maximizes may not be true; certainly there seems no reason why it should necessarily be true.
(c) The general presumption that revenue maximizers advertise more than profit maximizers is not necessarily valid either.

Figure 3 showed a case where the revenue maximizer produced a larger output and advertised more heavily than the profit maximizer. This fits what Baumol claimed would in general be true. But much depends on the responsiveness of costs and revenue to changes in output and advertising. There is no reason why the cost and revenue functions should not be as in Figure 4 below.

In Figure 4 the profit maximizer is shown at point X with output Q_1 and advertising of £100. The revenue maximizer with a constraint level of profits of £20 is shown at point Z with a lower output (Q_0) and higher advertising (£400). This is simply a case where in the search for maximum revenue for a given level of profits it is found that the best solution is to significantly raise advertising. Having shifted the demand curve upwards it is found that on the new demand curve it pays the revenue maximizer to charge a high price and sell a relatively low output. In other words, this is a case where the effect of advertising is such that it pays in terms of revenue to use the advertising to gain a higher price rather than to gain a greater quantity of sales. The generally held assumption that Baumol's model leads to

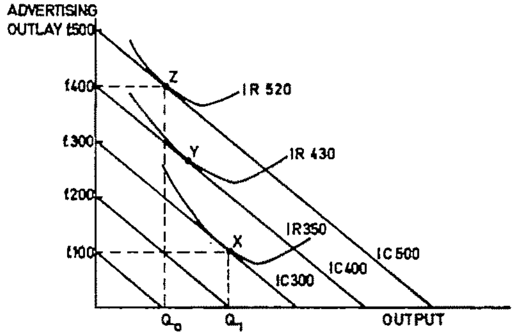


FIGURE 4

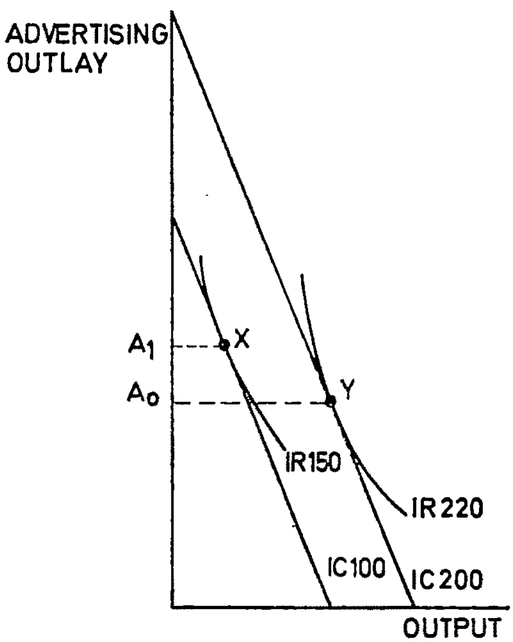


FIGURE 5

greater output than profit maximizing models does not therefore seem valid.

Neither, incidentally, is it necessarily true that revenue maximizers will advertise more heavily than profit maximizers. It is easily shown using a similar diagram to Figure 4 that revenue maximization can occur at lower levels of advertising than profit maximization. (See Figure 5.) The profit maximizer is shown at X with higher advertising outlay than the revenue maximizer at Y.

Finally, the equilibrium derived by R. L. Sandmeyer for the revenue maximizer with constant total costs can be shown to be correct, though the argument is not repeated here since it is covered by Kafoglis and Bushnell in the current issue of this *Review*.

II. *Conclusions*

First, an alternative to the Haveman and DeBartolo two-dimensional diagram can be devised to show the correct solution to the general Baumol model.

Second, using this alternative diagram it can be shown that revenue maximizers need not necessarily produce more output and advertise more than profit maximizers. They may produce less output (with higher advertising and a higher price). Equally they may advertise *less* than profit maximizers (with a higher output and a lower price).

It is stressed that all the above conclusions apply to single product firms. As I have shown elsewhere revenue maximization leads in general to the same results as profit maximization for multi-product firms.

REFERENCES

W. J. Baumol, *Business Behavior, Value and Growth*, rev. ed. New York 1967.

R. Haveman and G. De Bartolo, "The Revenue Maximization Oligopoly Model: Comment," *The Amer. Econ. Rev.* Dec. 1968, *58*, 1355–58.

C. J. Hawkins, "On the Sales Revenue Maximization Hypothesis," Southhampton Univ., Disc. Pap. in Economics and Econometrics No. 6806, 1969; *J. Ind. Econ.* forthcoming.

R. L. Sandmeyer, "Baumol's Sales Maximization Model: Comment," *Amer. Econ. Rev.*, *Dec.* 1964, *54*, 1073–81.

The Revenue Maximization Oligopoly Model: Reply

By ROBERT HAVEMAN AND GILBERT DEBARTOLO*

The comments of Milton Kafoglis and Robert Bushnell and C. J. Hawkins correctly describe an inconsistency in our earlier note due to an erroneous graphical representation of the "generalized Baumol-model" equilibrium.

The formal model presented in our note demonstrates that the revenue maximizer will devote profits in excess of the minimum to advertising so long as the marginal revenue per dollar of advertising exceeds the marginal revenue per dollar of production expense. In equilibrium, therefore, total revenue will equal total costs including production and advertising costs plus minimum acceptable profits. $[TR(A=a)=C(x)+a+\pi]$. In addition, as Kafoglis and Bushnell point out, the model also implies that dC/dx

* The authors are, respectively, professor of economics, Grinnell College, and graduate student in Economics, Massachusetts Institute of Technology. Mr. DeBartolo was an economics student at Grinnell College when our 1968 article and this reply were written.

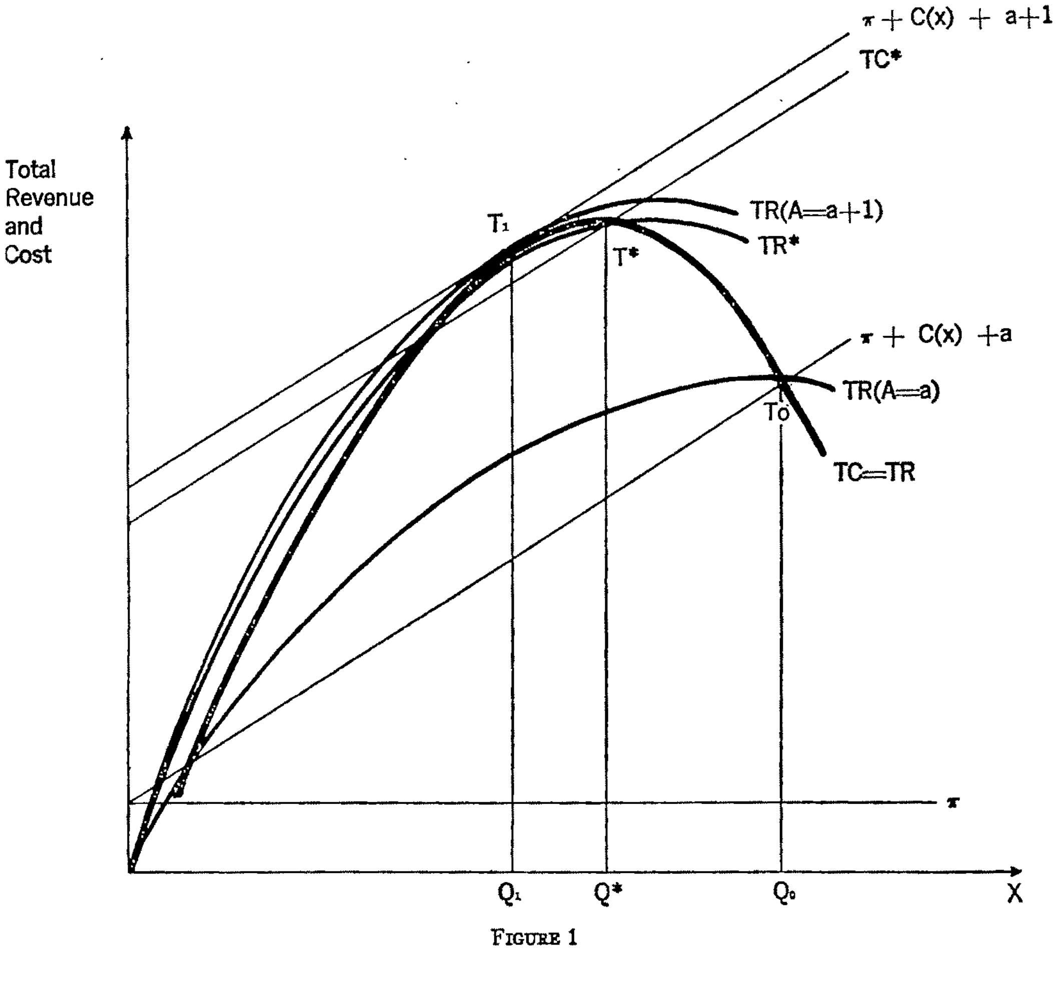


FIGURE 1

$>\partial TR/\partial x$ in equilibrium. However, the diagram in the note erroneously equates marginal cost and marginal revenue in equilibrium. It is, therefore, inconsistent with the model. Kafoglis and Bushnell accurately point out that the equilibrium must show a greater output, lower price, and less advertising expense than is implied in the diagram.

Figure 1 displays the correct diagram consistent with the model shown in our equations (1)–(5).[1] In Figure 1, T_0 is the equilibrium solution claimed by Robert Sandmeyer, which we demonstrated as impossible. T_1 is the original equilibrium solution we depicted which Kafoglis and Bushnell found to be inconsistent with the generalized Baumol model. The $TC=TR$ curve is the locus of points at which the total cost functions (including π) for various levels of advertising outlay intersect the total revenue functions generated by the corresponding advertising outlay. It represents all points at which $TR(A=a)=C(x)+a+\pi$.

[1] It should be noted that the sign preceding λ in (2) and (3) should be positive.

As we note in the first paragraph, the equilibrium must lie somewhere on this function. Although both T_0 and T_1 qualify as equilibria in that they both lie on this function, neither provides the maximum revenue attainable. As a revenue maximizer (and hence a total cost, including π, maximizer), the producer will find his equilibrium where the $TC=TR$ function attains a maximum. This is shown at T^*. At this equilibrium, the $dC/dx>\partial TR/\partial x$ equilibrium condition of the model holds. Thus, if the revenue maximizer was producing output Q_1, he would substitute production expense for advertising expense until output increased to Q^*. The lower price implied by the larger output increases total revenue by more than the reduction in advertising decreases it.

REFERENCES

R. Haveman and G. DeBartolo, "The Revenue Maximization Oligopoly Model: Comment," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1355–58.

R. L. Sandmeyer, "Baumol's Sales-Maximization Model: Comment," *Amer. Econ. Rev.*, Dec. 1964, *54*, 1073–1081.

Distributional Equality and Aggregate Utility: Comment

By WILLIAM BREIT AND WILLIAM P. CULBERTSON, JR.*

Although the starting point for modern discussions of the optimum division of income is the classic formulation of the problem given by A. P. Lerner, it is difficult to find an unambiguous statement of his analysis in the recent literature of welfare economics. Various interpreters, claiming equally the authority of Lerner, have offered differing and often contradictory statements of Lerner's argument and its implications. Lerner himself has noted this fact, remarking somewhat wistfully, that though he " . . . feel[s] very pleased . . . " with his argument it is nevertheless " . . . the least successful of my inventions."[1]

Our purpose in this paper is to show that the puzzlement and suspicion surrounding the Lerner theorem given in chapter 3 of *The Economics of Control* probably derives from his ambivalent statement of conclusions that allegedly follow from a single argument. For, as we shall see, there is not one Lerner theorem but two, and Lerner proves his case for only one conclusion (and here only under the most restrictive assumptions) and not the other. However the unsubstantiated Lerner conclusion can still be salvaged through a modified conceptualization of the problem.

In Section I, we shall summarize Lerner's statement of his theorem on income distribution under conditions of diminishing marginal utility of income and extend the model to accommodate the possibility of increasing marginal utility as well. In Section II, we shall review a recent revision of the theorem in terms of a voting model and suggest how the unanimity conclusion of this model can be preserved even with the existence of risk takers. In Section III, we shall indicate difficulties encountered when it is realized that Lerner's proof applies only to a mild version of his theorem and not to his bolder conclusion that allegedly follows from his argument. Section IV will be devoted to showing that Lerner's egalitarian conclusion can be rigorously vindicated leaving his basic strong theorem intact.

I. *On the Optimum Division of Income*

Mr. Lerner's Remarkable Theorem

Readers of The *Economics of Control* will recall Lerner's discovery that "If it is impossible, on any division of income, to discover which of any two inidviduals has a higher marginal utility of income, the probable value of total satisfactions is maximized by dividing income *evenly*" (p. 29). In arguing his case, Lerner relied upon the following five assumptions.

1. ". . . it is not meaningless to say that a satisfaction one individual gets is greater or less than a satisfaction enjoyed by somebody else" (p. 25). In other words, the utilities of different individuals are commensurable and addable.
2. "There is no way of discovering with certainty whether any individual's marginal utility of income is greater than, equal to or less than that of any other individual" (p. 28). That is, on any given distribution of income, a person's capacity to enjoy income is statistically independent of the amount of income he receives.
3. ". . . the principle of diminishing marginal utility of income holds generally" (p. 26).
4. ". . . satisfaction is derived only from one's own income" (p. 36). That is, utility functions do not exhibit interdependence.
5. The analysis is introduced as ". . . the

* The authors are, respectively, associate professor of economics and acting assistant professor of economics, University of Virginia. They are indebted to J. M. Buchanan, H. M. Hochman, and Ivan Johnson for helpful comments on an earlier draft of this paper.

[1] Personal letter from A. P. Lerner.

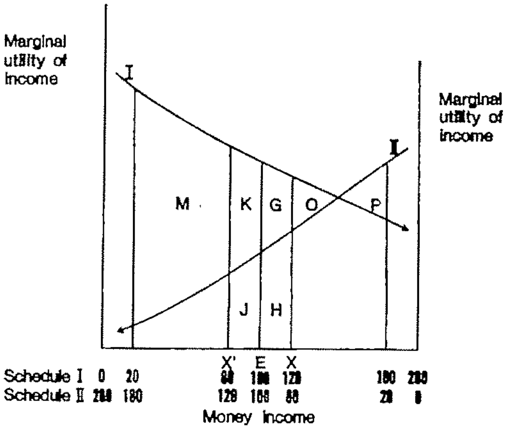


FIGURE 1

solution to the problem of maximizing the probable total of satisfactions that can be attained by the member of society from a given income" (p. 40). Thus Lerner assumes that the total amount of money income available is not affected by its distribution.

Let us refer to Figure 1 for an illustration of this argument.[3] By assumption (3) we know that the marginal utility of income schedules correspond to those illustrated in Figure 1. But by assumption (2) we cannot know to which individual to assign each schedule. Thus we have no way of knowing that division of income which will maximize total satisfactions. In effect then, we know that starting from any unequal division of income and given a movement to an equal (or more equal) division of income, we are just as likely to have diminished total satisfactions as to have increased them. For example, in Figure 1 suppose person A to have a money income of \$120 and person B to have a money income of \$80. This unequal division of income may correspond to point X in Figure 1 where Schedule I will be that of person A and Schedule II that of person B. If we were now to transfer \$20 from A to B (i.e., in Figure 1 a movement from point X to point E) we would increase B's satisfactions by area H but diminish A's satisfactions by area H plus area G. There would be a net loss of satisfactions equal to area G. But it is equally likely by assumption (2), that the initial division of income ($A=\$120$, $B=\$80$) corresponds to point X' where Schedule II is that of A and Schedule I that of B. Under these equally likely circumstances, the transfer of \$20 from A to B (i.e., a movement from point X' to E in Figure 1) would diminish A's satisfactions by area J but increase B's satisfactions by area J plus area K. There would be a net gain in satisfactions equal to area K. Since either result (net loss of G or net gain of K) is equally likely and since the gain exceeds the net loss ($K>G$) the maximization of the probable total value of satisfactions requires that income be divided evenly.

Note that this conclusion holds irrespective of the initial degree of inequality. Should A have originally possessed \$180 and B, \$20, a movement to an equal division of income is just as likely to produce a net gain in total satisfactions (area M plus area K in Figure 1) as it is to produce a net loss in total satisfactions (area G plus area O less area P). Since once again the net gain exceeds the net loss, it follows that the maximization of the probable total value of satisfactions is realized by an egalitarian division of income.

A Generalization of the Lerner Theorem

While Lerner and his commentators, M. Friedman (p. 308), I. M. D. Little (p. 60), R. Musgrave (p. 106), P. A. Samuelson (p. 173) feel that the assumption of diminishing marginal utility of income to all individuals is needed to make the case for equality, it is not a necessary assumption in all circum-

[3] For facility of exposition Figure 1 presents a slightly modified version of Lerner's diagram which appears on page 30 of the original. Lerner preferred to demonstrate his proposition by comparing the resulting gains and losses due to a movement from an equal to an unequal division of income (that is, a movement from point E to either point X or point X' in our Figure 1). This approach has misled at least one writer: B. de Jouvenel asserts that Lerner's demonstration ". . . rests upon the highly artificial assumptions that the initial condition is one of equality and that moves away from it are haphazard" (p. 33). In order to dispel the notion that Lerner's argument is an argument for the status quo distribution of income (however unequal that distribution might be) we prefer to begin the analysis with income distributed unequally. In doing so it seems felicitous to number the utility schedules as we have done in Figure 1.

stances. Consider the possibility (in Figure 2) of either (but not both) person A or person B (we cannot know which one) having an increasing marginal utility of income schedule. Further suppose the relevant schedules to be Schedule I (diminishing marginal utility of income) and Schedule II (increasing marginal utility of income) and assume A to have \$120 and B, \$80. If we now divide income evenly, it is just as likely that there will be a net gain (area K in Figure 2) as it is that there will be a net loss (area G).[3] Since the equally likely net gain will exceed the net loss, it is still the case that the maximization of the probable total value of satisfactions is realized by dividing income equally.

However, such a conclusion cannot be reached if Schedule II is replaced, in Figure 2, with Schedule II′—where the rate of increase of Schedule II′ exceeds the rate of decrease of Schedule I for the relevant income levels. Given the same initial degree of inequality ($A = \$120$, $B = \$80$) a movement to an equalitarian distribution of income is just as likely to produce a net loss (area Q) as it is to produce a net gain (area R). Since the equally likely loss exceeds the gain it does not follow that in these circumstances the maximization of the probable total value of satisfactions is achieved by dividing income evenly.

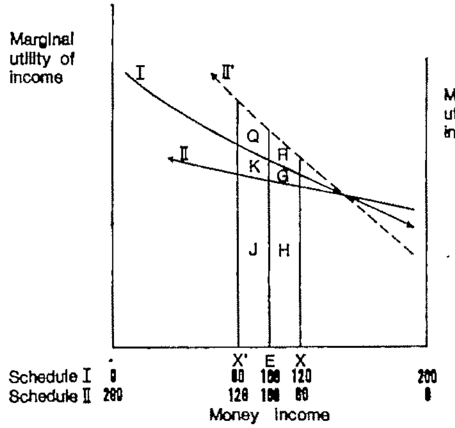


FIGURE 2

II. *A Voting Model Reformulation*

What is remarkable about Lerner's theorem, of course, is that starting from value-free postulates it arrives at a seemingly ethical judgment about income distribution. This apparently paradoxical result has led many writers to view the Lerner theorem with great suspicion.[4] Lerner appears to have performed the illusionist's feat of producing something out of nothing. His puzzled onlookers, in attempting to discover the secret of his magic, have focused their attention on the seemingly innocuous assumption of "equal ignorance."[5] Thinking they have exposed the source of Lerner's sleight-of-hand, they have insisted that the equal ignorance premise be modified.

Samuelson, for example, develops a model that is "spiritually akin to Lerner's" but which shifts the equal ignorance assumption to a different level. Samuelson shows that the Lerner equality rule would be voted in unanimously at the constitutional level of choice in a world of risk averters (i.e., if all individuals have downward sloping marginal utility of income schedules). If we assume that each individual is equally ignorant about the income that he will receive in the future,[6] then from the perspective of each individual it is equally likely that he will receive *any* income dealt out *randomly* by fate. Imagine that total social income is divided up into as many magnitudes as there are individuals in our society and each

[3] Following the analysis given earlier, the redistribution may proceed from point X (in Fig. 2) where Schedule I is that of A and Schedule II is that of B. In this case B gains H while A loses $H+G$, with the resulting net loss of G. But it is equally likely that redistribution will correspond to a movement from point X' to point E, where Schedule II is that of A and Schedule I is that of B. Thus A loses area J while B gains area J plus area K, with the resulting net gain of K being greater than the equally likely net loss of G.

[4] Paul Samuelson, celebrating Lerner's sixtieth birthday, wrote that he felt doubtful, "... that you can reap ethical results without planting ethical postulates" (p. 176).

[5] See assumption (2). The term "equal ignorance" as used in this context seems to have originated with Friedman (p. 308).

[6] This assumption has its counterpart in Lerner's explicit statement that his model applies only to inequality caused by chance or inheritance and not by differences in services rendered.

magnitude is written on a separate card. If the cards are then shuffled and dealt out at random, each individual can expect with equal probability to receive any card. If he is a risk-averter he would prefer that income be divided evenly rather than take a chance on getting too low an income. So a world of risk-averters would prudently and unanimously vote for equality. Note that in this voting model version of Lerner's theorem equal ignorance no longer means that the marginal utility of income schedule of one man is equally likely to be the same as, higher than or lower than that of another. Now it means that every permutation of incomes that an individual can expect to receive is equally likely. So Samuelson does not jettison the equal ignorance assumption but only substitutes one kind of equal ignorance assumption for another.

Since Samuelson argues that the existence of risk lovers in this society would keep the vote from being unanimous, this might seem inconsistent with our earlier argument that the existence of diminishing marginal utility of income schedules for all individuals is not necessary for Lerner's equality rule to hold. But by a rather simple extension this voting model approach can be made consistent with Lerner's egalitarian conclusion even in the case in which some individuals have increasing marginal utility of income schedules.

Consider the case in which individuals with a relatively high intensity for risk aversion (represented in Figure 2 by Schedule I) face individuals with a relatively low but nevertheless positive preference for risk (represented in Figure 2 by Schedule II). Although the marginal utility of income for some members of society is increasing there is still the possibility of a unanimous egalitarian vote. For imagine once again that society's income is to be dealt out at random to each individual who is equally likely to receive an income of any magnitude. If a unanimity rule is required for equality, it might appear that the existence of one risk lover would keep the division of income unequal. But this overlooks the possibility of vote-trading through logrolling. (See J. M. Buchanan and G. Tullock, pp. 125–26.) If the intensity of preference of the individuals represented by Schedule II against equality is less than the desire of other voters in favor of equality (represented by Schedule I), the majority who would be the potential gainers would compensate the risk lovers (potential losers) by selling their votes on issues that are of more intense relative interest to the latter. In this way an egalitarian division of income would be unanimously accepted at the constitutional level of choice even in the face of increasing marginal utility of income functions for some persons.

III. *Some Ambiguities in the Lerner Theorem*

Not the least of the reasons for the dissatisfaction with the Lerner theorem is that Lerner's argument is itself deeply ambiguous. For a careful reading will indicate that Lerner states not one but two propositions in chapter 3 of his book. One of them is the relatively meek and mild assertion that "... the maximization of *probable* total satisfaction is attained by an equal division of income" (p. 29). The other is the much bolder claim "... *that if it is desired to maximize the total satisfaction in a society, the rational procedure is to divide income on an equalitarian basis*" (p. 32, Lerner's italics). Notwithstanding J. E. Meade's statement that Lerner has offered an interesting and elegant proof (p. 49) of the relatively meek proposition, everyone, including Meade, has remained silent on the bolder proposition, not realizing perhaps that Lerner has switched his conclusions between page 29 and page 32. There is a subtle difficulty in the Lerner formulation that has gone unnoticed when the relatively mild proposition regarding probable satisfactions is switched to the much bolder and more interesting proposition regarding realized satisfactions.[7] But this difficulty is sufficient in itself to in-

[7] We shall generously interpret Lerner's strong theorem to be that if it is desired to *increase* the total satisfaction in society, income should be divided on an egalitarian basis. For it should be clear that Lerner cannot literally mean, under his assumption of equal ignorance about utility functions, that an egalitarian division of income insures that the marginal utility of income for each individual is the same.

capacitate the proof. In short, Lerner has proved only his weak proposition for a two-person model while asserting his bolder one for an N-person society.

What we mean is perhaps best explained by considering the following simple model, which in essence contains all of the assumptions of Lerner's two-person example. Two men, A and B, inhabiting an island with a direct telephone connection to Professor Lerner, have the problem of dividing a given supply of coconuts. Suppose they call Lerner to ask him for advice and he replies in conformity with his famous chapter 3, "If it is desired to maximize the probable total satisfactions on your island, the rational procedure is to divide the coconuts on an equalitarian basis."

Having done this they might have a vague feeling of uneasiness, and before agreeing to this equal division as final, they call Lerner once more. They tell him, "Professor Lerner, can you reassure us that, having followed your advice, we are probably better off than with an unequal division of coconuts?"

Notwithstanding his statement of the strong proposition of page 32, Lerner would *not* be in a position to reassure them. He could only reply, "Since I am equally ignorant of your respective utility functions, I can only say that *if* you have made yourselves better off, you have increased your gains in satisfactions more than *if* you have made yourselves worse off. But I cannot say that you are better off or even probably better off. You are just as likely to be better off as worse off." Since this is so, Lerner cannot, under his assumptions, ever assert his strong proposition that the men will increase their satisfactions by following his advice on equal distribution.

But Lerner might reply that he did not mean for his argument to apply in redistribution between *only* two individuals. He could claim that probabilities enter his argument in a significant way when he shifts from a two-person model to one containing a large population. As he put it, "out of 100 million shifts away from an equalitarian distribution of income in a large population, it could be expected that about 50 million would increase total satisfaction and about 50 million would diminish it. In about 50 million cases the shift would be beneficial . . . and in the other 50 million cases it would be harmful . . ." (p. 32). In the aggregate, however, the losses would be greater than the gains. But as we shall show, even in this aggregate case, he cannot assert this bold claim.[8] To see that this is so, let us translate his example into an aggregate case of our island model.

Consider 100 million islands of the sort described above, each containing the same number of coconuts and inhabited by two men, A and B, with direct communications to Lerner. If Lerner were to receive a call from each island asking him how to divide the coconuts, he would say, "Divide evenly." But this advice will certainly not do in the aggregate. For now he has no way of knowing that in the 50 percent of those instances when gains were made, they were greater than the losses in those 50 percent of the instances when losses were made. For without any prior knowledge about the utility functions of the 200 million individuals he cannot compare the areas under the marginal utility curves of the different individuals on the different islands. Once again he can say that for *each* island if gains were made, they would be greater than the losses if losses were made. But this does *not* mean that in the aggregate the gains that were made exceed the losses. Thus there is no assurance that an aggregate gain will be realized by an egalitarian division of income.

The failure of Lerner's strong theorem is not a failure in its basic conclusion, i.e., that with equal ignorance about utility functions total satisfaction will increase by dividing income equally. It is rather a failure of proof stemming from an incorrect conception of the problem. However, a slightly different con-

[8] Indeed, in the aggregate case his argument breaks down even for his weak proposition unless Lerner explicitly assumes that the same degree of inequality holds for all individuals throughout society. In short, Lerner must imagine a strangely Orwellian world in which all men are unequal but none is more unequal than others.

ceptualization and the addition of one assumption will suffice to rigorously prove even Lerner's strong theorem (which of course subsumes the weak theorem) and put discussion of it on a fruitful basis.

IV. *A Rigorous Pooof of Lerner's Theorem*

"The gods link like with like."
Homer, *Odyssey*

Let us drop the equal ignorance assumption for the moment in order to allow Lerner to inspect the utility functions of all individuals in our society. He can now choose any individual at random (e.g., person A with, say, income of \$120 and with utility function Z) and place him on one of our islands. Now, on the assumption that every man has his identical twin in capacity to enjoy income, Lerner chooses A's twin (person B) and puts him on the island along with A. That is, both A and B have utility function Z.

But it is clear that A and B will not have equal income since, according to assumption (2), income is statistically unrelated to a person's capacity to enjoy it. Therefore, let us assume that B's income is any amount different from A's, say \$80. However the mean income of A and B will be equal to the mean income of society. Now let Lerner pick a third person, C, who has an *income* equal to that of A (i.e., \$120) but with a necessarily different (by assumption (2)) utility function (e.g., utility function X). Then pick another person, D, who also has utility function X, that is, he is C's identical twin in utility. Once again by assumption (2) this necessarily implies that person D has an income different from that of person C, but that the mean income of C and D is equal to the mean income of society. Thus D's income is \$80. And so we assume, along with Homer, that Lerner, like the gods, pairs off like with like throughout society in a fashion similar to that described above. Once this has been done, he can again invoke his equal ignorance assumption. For, he does not need to know whether the utility functions of persons A and B are greater than, equal to, or less than the utility functions of persons C and D. Nor does he need to assume that in the aggregate the same degree of inequality holds for all individuals. That is, the initial distribution of income among the persons on each island can be quite different from that on any other island. Given this conceptual classification could Lerner now assure the denizens of each island that, once they have divided income equally, they will have maximized probable satisfactions (the weak Lerner theorem) *and* achieved an increase in actual satisfactions? Lerner could now assure them that this would indeed be the case. For in that circumstance in which the movement to equality was wrong (i.e., reduced satisfactions) the loss would be more than offset by the gain that was made in the case in which the movement toward equality increased satisfactions. Therefore the total satisfactions on each island would be greater, and since the whole is equal to the sum of its parts, the total satisfaction in the aggregate society of islands would be greater by following the advice of an egalitarian division of income. Furthermore, since the per capita income on each island is the same, equal distribution within each island means equal distribution throughout the aggregate of islands. Since we can view the aggregate of these islands as the society, Lerner's strong conclusion has been rigorously derived.

Furthermore, it is possible to extend the argument to include the case of upward sloping marginal utility of income functions, so long as the pairs of schedules are always conceptually matched as are Schedules I and II in Figure 2. Once this conceptual experiment is conducted, the general problem of the optimum division of income, given Lerner's assumptions *plus* the crucial assumption that every individual has his identical twin in capacity to enjoy income, is solved.[9]

[9] While Friedman attempted to make Lerner's proof rigorous through a conceptual experiment in his review article (p. 309), he has convinced some that he has "produced a mouse and not Lerner's mouse" (see Samuelson, p. 176). Perhaps the reason that Friedman's restatement has not been generally accepted is that: (a) He does not make a clear distinction between the weak and strong Lerner theorems which we believe to be a fertile source of confusion in this issue. (b) In Friedman's schema each individual faces another with pre-

REFERENCES

J. M. Buchanan and G. Tullock, *The Calculus of Consent.* Ann Arbor 1962.

M. Friedman, *Essays in Positive Economics.* Chicago 1953.

B. de Jouvenel, *The Ethics of Redistribution.* Cambridge 1952.

A. P. Lerner, *The Economics of Control.* New York 1944.

I. M. D. Little, *A Critique of Welfare Economics.* London 1950.

J. E. Meade, "Mr. Lerner on 'The Economics of Control,'" *Econ. J.*, Apr. 1945, *55*, 47–69.

R. Musgrave, *The Theory of Public Finance.* New York 1959.

P. A. Samuelson, "A. P. Lerner At Sixty," *Rev. Econ Stud.*, June 1964, *31*, 169–78.

---

cisely the same utility function in the redistributive process. This seems outside the spirit of Lerner in that Lerner's equal ignorance assumption allows for the possibility that redistribution may take place between individuals with dissimilar utility functions. (c) Friedman's formulation of the conceptual experiment leaves unresolved the problem that there may be some individuals with unique utility functions. To make Lerner's thorem rigorous under this possibility, Friedman must have the implicit assumption that in such cases income must have already been divided equally.

In addition to avoiding the above pitfalls, our analysis offers a rigorous proof of a more general Lerner theorem in that if allows us to accommodate, under certain circumstances, the existence of increasing marginal utility of income functions. Friedman's proof would be incompatible with this possibility.

Distributional Equality and Aggregate Utility: Reply

By A. P. LERNER*

The formulation of the argument for distributional equality by William Breit and William Culbertson is an improvement on that of *The Economics of Control* and is much more effective in class. Their generalization of the proposition to the case of increasing marginal utility (of income) offset by a greater degree of diminishing marginal utility elsewhere, is also an improvement. Their point that Paul Samuelson did not escape the "equal ignorance" assumption is well taken.

Ambiguity, being a case of lack of clarity is a charge that can never successfully be refuted. However, I would like to deny a switching of conclusions. Perhaps the ambiguity would have been avoided if I had added the following words in Roman type to the italicized sentence quoted: ". . . *if it is desired to maximize the total satisfaction in a society, the rational procedure*, in the absence of the knowledge that would enable us to equalize the marginal utilities, *is* to maximize the probable total satisfaction—i.e., *to divide income on an equlitarian basis.*" The theorem on page 32 is not "bolder" than the "meek and mild" one of page 29. It is the same proposition.

The ingenous device of the 100 million coconut islands in one way does more than is claimed for it and in another way, does less. If it were possible to divide the total population into pairs which had the same utility functions, the equalization of income within each pair would *never* involve a "wrong" movement to be offset by a "right" one. That is why there is certainty of improvement from equalization on every island. Furthermore, there would be an absolute maximization, with certainty, of the total satisfaction of the pair on each island from their joint income.

On the other hand, the parable assumes that the combined incomes of the pairs have somehow already been equalized; that for every individual in the half of the total population with incomes less than the mean, his "partner" in the other half of the population (with an identical utility function) has an income greater than the mean by the exact amount that his is *less* than the mean. (This implies incidentally that no individual has an income as much as twice the mean unless his "partner" has a zero income.)

If this is not the case, some islands will be richer than others. We will then have to equalize the incomes of the islands before we could conduct Breit and Culbertson's experiment. The parable, therefore, while not necessary for the "meek" proposition that income equalization maximizes the *probable* total satisfaction, is not sufficient for the "bold "proposition (to which I have never subscribed) that income equalization increases total satisfaction with *absolute* certainty.

Breit and Culbertson's development of their parable reflects the same discomfort they have seen in others. The pair on the island are not satisfied with the proof that the equalization of the incomes has *maximized their probable* satisfaction. Sharing a widespread human craving for certainty, they want to be *quite sure* that they have at least *increased their actual* total satisfactions. This assurance is unfortunately not available as long as the utility functions are unknown.

Breit and Culbertson also are seeking for a certainty of gain in a "much bolder and more interesting proposition regarding realized satisfactions" instead of the maximization of a mere probability, and are accurately represented by the island pair they have invented. They have imagined a certainty of gain only by imagining the discovery of identical-utility twins. But the whole point of the

* University of California, Berkeley.

argument is that in the absence of knowledge of the utility functions (so that we cannot equalize the marginal utilities—with different incomes for individuals with different utility functions and equal incomes for those, like the island pairs, with identical utility functions) we nevertheless still have the probability proposition.

There can be no absolute certainty of gain from any redistribution between individuals of whose utility functions we are ignorant. Nevertheless, if I were offered 11 cents for every "head" in return for 10 cents for every "tail" on 100 million tosses of an unbiased coin, I would consider the probability of gain certain enough.

Allais' Restatement of the Quantity Theory: Comment

By MICHAEL R. DARBY*

In a recent article in this *Review*, Maurice Allais proposed a model of the demand for money based on "the 'psychological rate of expansion'" (p. 1129). His formulation of the model has certain arithmetic implications that appear implausible.

The psychological rate of expansion (Allais' z) is comparable to Phillip Cagan's "expected rate of change in prices" (Cagan's E, p. 35).[1] Although Allais, like Cagan, relates his psychological rate of expansion to remembrance of past events, he is not specific about its precise economic meaning (pp. 1129–30, 1155).[2] In particular, there is some confusion whether his variable z or $Z=250z$ is the appropriate counterpart to the expected rate of change in prices.[3] However, it seems clear, despite some effort by Allais to differentiate his product, that z, an estimate of anticipations of growth in total outlays, is his basic construct while Z is introduced "[t]o facilitate the econometric analysis. . . ." (p. 1132).

Allais' equation (2.28) defines Z by

$$(1) \qquad Z(t) = \frac{z(t)}{\chi_0},$$

and his equation (2.50) sets $\chi_0=0.004$; so

$$(2) \qquad Z(t) = 250z(t)$$

Thus

$$(3) \qquad \frac{dZ}{dt} = 250\frac{dz}{dt}$$

The fundamental differential equation (2.52) relating dZ/dt to dx/dt is

$$(4) \qquad \frac{dZ}{dt} = x - 0.002(1 + e^{Z})Z$$

Alternatively, we can write

$$(5) \qquad \frac{dz}{dt} = x - 0.5(1 + e^{250z})z$$

Thus z is adjusted by the amount of the deviation of experienced expansion x from a function of anticipated expansion z. Either of these equivalent equations can be used to define unique equilibrium values of z and Z ($\equiv 250z$) corresponding to an indefinitely constant value of x. To do so, set

$$(6) \qquad \frac{dz}{dt} = \frac{dZ}{dt} = \frac{dx}{dt} = 0$$

This gives

$$(7) \quad x = .002(1 + e^{Z})Z = .5(1 + e^{250z})z$$

Choosing various values of z, we can determine corresponding values of x for which each value of z is an equilibrium value. These values of z and x are given in Table 1 to-

* The author wrote this note while a National Science Foundation Graduate Fellow at the University of Chicago. He wishes to thank Professor Milton Friedman for help in clarifying several points in the exposition.

[1] Using Allais' comment and equation (2.11) (p. 1130) and his fn. 2 (p. 1123).

[2] Allais' psychological time scale presents the main difficulty. Under his postulates, if output expanded at a constant positive rate forever, psychological time would pass faster than chronological time. Hence x, the rate of growth of outlays per month, would exceed x', the rate of growth of outlays in terms of psychological months. Using equations (2.9) and (2.11), we have $z=z'=x'$; so that $z\neq x$. Thus Allais' psychological rate of expansion ($z'=z$ in this case) differs from the totality of past experience. This difference, $x-z$, increases as the constant positive x is increased. (See Table 1 and the discussion of its formulation, below.) It seems to me to be a defect of Allais' analysis that, regardless of all past experience, the psychological and chronological time scales would be identical only for a zero rate of expansion.

[3] Note Charts 7 and 8 (pp. 1147–8) and z' defined as "the 'psychological rate of expansion' " (p. 1129) and "... a rate Z, designated as the psychological rate of expansion. . . . " (p. 1151).

gether with $\varphi(z) = \psi(Z)$, the relative demand for cash balances.[4]

Table 1 casts considerable doubt upon Allais' interpretation of his model. Presumably, a forever constant rate of experienced expansion would lead to expectations (or, in Allais' terms, memories) of the same rate of expansion. However Table 1 makes clear that Allais' equation (2.52) does not have this implication.[5] Indeed, expectations are so sluggish that an expansion of 13 percent per month would lead to expectations of only a little over 1 percent per month (z for $x=0.1265$ is 0.012.); even more incredibly, an expansion of 44,054 percent per month, maintained indefinitely, would lead to expectations of only 4 percent per month. Similarly, the implied demand curve for money balances is hard to accept. It indicates that an expectation that prices (or outlays) will increase by 4 percent per month ($z=.04$) induces holders of cash balances to hold only nine-one hundred thousandths of a dollar for each dollar they would hold if they expected a zero price increase ($\varphi(z)$ for $z=.040$ is .00009).

An additional difficulty is that the model is constructed in terms of outlays[6] or, in the empirical tests, nominal income.[7] Yet, it is irrational to hold money balances equal to the same percentage of nominal income when prices are rising 5 percent per year and output is constant as when prices are constant and output is growing 5 percent per year. The rate of inflation is the "own price" of holding money and it is hard to see why the rate of growth of real income should have an identical effect.[8]

Comparison of the values of x and $\varphi(z)$ in

TABLE 1—EQUILIBRIUM VALUES OF z & $\phi(z)$, GIVEN x[a]

z	x	$\phi(z)$	Z
0	0	1	0
0.0008	0.0009	0.90033	0.2
0.0016	0.0020	0.80263	0.4
0.0024	0.0034	0.70869	0.6
0.0032	0.0052	0.62006	0.8
0.0040	0.0074	0.53788	1.0
0.0080	0.0336	0.23840	2.0
0.0120	0.1265	0.09485	3.0
0.0160	0.4448	0.03597	4.0
0.0220	1.4941	0.01339	5.0
0.0240	4.8532	0.00495	6.0
0.0280	15.3664	0.00182	7.0
0.0320	47.7120	0.00067	8.0
0.0360	145.8738	0.00025	9.0
0.0400	440.5476	0.00009	10.0

[a] Note that although z was the variable chosen for computational purposes in the table, economically the independent variable is x and the dependent variables z and $\phi(z)$. Values of Z are appended to demonstrate, upon comparison with the values of x, the economic irrelevance of Z as a measure of the rate of expansion.

Table 1 indicates a not implausible relationship. The elasticity of $\varphi(z)$ with respect to x is on the order of $-.3$ for moderate inflations and approaches -1.0 for more rapid inflations. This illustrates the combined effect of Allais' equations (2.51) and (2.52); the first defining a demand curve for money, the second, the formation of expectations. The extremely inelastic expectations model is precisely offset by an equally elastic demand for money function. Allais has solved the problem of separating χ_0 and α by use of postulates instead of a fitting method following Cagan. $1+e^Z$ in the numerator of (2.52) is offset by the $1+e^Z$ in the denominator of (2.51).

Despite the difficulties with Allais' precise formulation, the breakthrough in the functional form is truly significant. Allais' basic expectational equation is (2.27):

$$\frac{dz}{dt} = \chi_0\left[x - \frac{z}{\varphi(z)}\right] \tag{8}$$

[4] $\phi(z_0)$ is the ratio of (1) desired nominal holdings of money divided by nominal outlays when $z=z_0$ to (2) desired nominal holdings of money divided by nominal outlays when $z=0$.

[5] See fn. 2.

[6] Equation (2.7).

[7] Equation (3.7). In hyperinflations, however, Allais uses the growth rate of prices for x in equation (3.10).

[8] To the extent permanent or life-cycle income is a motive for holding cash balances, expected growth in real income should raise, not lower, the relative demand for cash balances. Allais' adjustment of ϕ_0 and k for each country would appear to compensate for different rates of growth in real income as well as taste variables.

Simple algebra transforms this to

$$
\frac{dz}{dt} = \frac{\chi_0}{\varphi(z)} \left[\varphi(z)(x - z) + (1 - \varphi(z))(k - z) \right] \tag{9}
$$

where $k = 0$.

Thus anticipations are adjusted according to a forgetfulness coefficient $(\chi_0/\varphi(z))$ times a weighted average of the deviation of the actual from the anticipated rate of change and the deviation of the long-run "normal"[9] from the anticipated rate of change where the weights are based on the ratio of psychological to chronological time. The first comment is that k should not be set equal to zero but allowed to vary through some very long period anticipations model.[10] The second is that $\varphi(z)$ is being worked very hard: It serves as the demand function for money, the adjustment in the forgetfulness coefficient, and the index of relative psychological time. At least for the latter purpose it is inherently defective: things would appear to be happening faster if prices were falling at 50 percent per month instead of remaining constant, while Allais' formulation (Chart 1, p. 1134) has psychological time slowing in this situation.[11] It would certainly seem appropriate to separate the demand function from the psychological time function because of the general implausibility that these two functions should be equal. A suitable psychological time function might be[12]

$$
\phi(z) = \frac{1 + b}{1 + be^{a|z|}}
$$

This reformulation appears to have great merit in explaining the formation of anticipations.

In summary, the particular form of Allais' model has highly questionable implications, but a reformulation of his anticipations model seems very promising.

REFERENCES

M. Allais, "A Restatement of the Quantity Theory of Money," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1123–57.

P. Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956, pp. 25–117.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

[9] The model is similar to that of Cagan with the addition of a concept of a normal rate of change and psychological time (or deviation from normal) adjustments.

[10] Such a model would be some average of, say, the last one hundred years. This allows z to adjust to any x maintained for long enough. This reformulation of Allais' model is due to Milton Friedman.

[11] While this is of minor concern in Allais' problem—we do not observe many "hyperdeflations"—it is of potential importance in other applications of the anticipations model.

[12] Cagan's demand function might be suitable since it does not have the problem of a maximum demand for real cash balances in a deflation. The z should be based on price, not income, changes.

Allais' Restatement of the Quantity Theory: Reply

By MAURICE ALLAIS

> The theories that summarize the state of our scientific knowledge are no doubt necessary vehicles for that knowledge. They should also provide support for further investigations of new ideas. But neither the theories nor the ideas are immutable: we must always be prepared to abandon, modify or replace them as soon as they no longer represent reality properly. In a word, theory must be adapted to fit nature: nature cannot be adapted to fit theory. . . .
>
> It follows that when a scientific idea has been put forward, the aim should not be to preserve it by seeking support for it at all cost, while discarding everything that might tend to invalidate it. The opposite should govern our behavior. The facts that seem to upset a theory should be examined with the greatest care, for real progress always consists of relinquishing an earlier theory which covers fewer facts by another of more extensive scope.
>
> Claude Bernard
> Introduction à l'Etude de
> la Médecine Expérimentale, 1865.

I am grateful to Michael R. Darby to have given me the occasion, with his interesting questions, to comment and render explicit essential points of the theory of the demand for money which I have proposed. This was completely impossible to do in the framework of the paper published in this *Review*.[1]

In his comment, Darby raises a considerable number of issues, but perhaps their presentation is quite confused. For clarity's sake, it is necessary to distinguish them clearly.[2]

Generally we have

$$(1) \qquad \frac{dZ}{dt} = x - \frac{\chi_0}{\psi(Z)} Z \qquad [2.43]$$

where x is the percentage rate of increase of global expenditure, relation [2.7].

When related to the psychological time, the corresponding relation is

$$(2) \qquad \frac{dz'}{dt'} = \chi'(x' - z') \qquad [2.10]$$

In a dynamic equilibrium process Z remains constant. Thus according to (1)

$$(3) \qquad Z = \frac{\psi(Z)}{\chi} x = \frac{x}{\chi}$$

so that,

$$(4) \qquad \chi Z = x$$

For such a process we have likewise from (2)

$$(5) \qquad z' = x'$$

Darby states that my definitional equations are inconsistent. In fact, his argument is inconsistent. From [2.2] and [2.8] of my *Review* article,

$$(6) \qquad x' = \frac{\chi'}{\chi} x$$

or, using [2.22] and [2.41]

$$(7) \qquad x' = \psi(Z)x$$

or again, from [2.51]

$$(8) \qquad x' = \frac{2}{1 + e^{z}} x$$

[1] For lack of sufficient space my article was too condensed to be easily intelligible. Thus to go more thoroughly into the new theory, it is necessary to refer to the French version which is quite detailed, (Paris 1965).

[2] For the same reason, the reference number following each equation is that of the corresponding equation in my *Review* article. When I refer to equations from the review article, the numbers will be in brackets.

and not $x'=x$ as Darby erroneously concludes. The point is that x' relates to the psychological time scale, x to the physical time scale.

It immediately follows that since *by hypothesis* $z=z'$, relation [2.11], and since *in a dynamic equilibrium process* $z'=x'$ according to (5), we have $z=x'$ *and not* $z=x$ as Darby erroneously states.[3]

Darby's error is to search for interpetations of the different parameters of my theory which accord with the theory of the expectation coefficients developed by Philip Cagan and Milton Friedman. This is an impossible task.

Thus *in a dynamic equilibrium process* in which x, x', and Z are constant

$$(9) \qquad z = z' = x'$$

The correspondence between x and Z is defined from (4) and thus according to the relation (3) above and the relation [2.51] of my *Review* article

$$(10) \qquad x = 0.002(1 + e^{z})Z$$

(see my "Growth and Inflation," Appendix B, § III) and, naturally, from (8), (9) and (10) there follows

$$(11) \qquad x' = \chi_0 Z = 0.004Z$$

Equations (10) and (11) define the correspondence between x, x', and Z *for a dynamic equilibrium process.*

Darby's table is correctly compiled. Preferably it should have been presented as in my Table 1 which shows clearly the relationships between the various parameters for a dynamic equilibrium process and their range of variation.

Darby's Table 1 is correct, but his comments on it are incorrect. He assumes that in relation to physical time, the parameter that plays the role of Cagan's coefficient of expectation is the parameter $z'=z=\chi_0 Z$. In fact, *this comparison is possible only when made in terms of the psychological time scale,* when the rate of forgetfulness χ is equal to χ_0. (See my *Review* article, p. 1130, first paragraph.)

But Cagan's expectation coefficient relates to physical time, and when this is used as reference scale, Cagan's coefficient has *no general equivalent* in my thoery.

Yet *in a dynamic equilibrium process*, we can easily see that the product χZ plays a role equivalent to that of Cagan's expectation coefficient. From (1) above and [2.41] we derive

$$(12) \qquad \frac{d(\chi Z)}{dt} = \chi[x - (\chi Z)] + Z\frac{d\chi}{dt}$$

and since *in a dynamic equilibrium process* we have

$$(13) \qquad d\chi/dt = 0$$

there follows

$$(14) \qquad \frac{d(\chi Z)}{dt} = \chi[x - (\chi Z)]$$

This relation shows that *for such a process and using physical time as reference scale,* Cagan's expected rate of price change is comparable to the product χZ, itself equal to x according to (4), *but this comparison is valid only for a dynamic equilibrium process and generally it does not hold.*

Thus we see that *in a dynamic equilibrium process*, Cagan's expected rate of price change can be compared with the product χZ, since these two quantities verify the same differential equation. Since in such a process and according to (4), the product χZ is equal to x, Cagan's expectation coefficient can be seen as equal to the rate of expansion x of global expenditure, and this is a completely natural result. In particular, a rate of expansion of 13 percent per month leads to Cagan's coefficient of expectation of 13 percent, and not 1 percent per month as Darby contends.

It follows that *in a dynamic equilibrium process* it is the product χZ, and not z or Z, whose role should be considered as analogous to that of Cagan's expected rate of change of prices. No ambiguity is possible.[4]

[3] See my *Review* article, p. 1130, first paragraph. A detailed comparison of the mathematically equivalent formulations, as far as the rate of psychological expansion is concerned, of Allais (1953–1954) and Cagan (1953–1954) is presented in Allais (1965), pp. 159–180.

[4] See below the comparison between the rates of interest and forgetfulness which makes the effective meaning of the products $\chi_0 z$ and χz easy to grasp.

TABLE 1—CORRESPONDENCE OF PARAMETERS IN A DYNAMIC EQUILIBRIUM PROCESS

Z^a	$\psi(Z)^b$	x^c	z^d	χ^e	ϵ^f
$-\infty$	2	$-\infty$	$-\infty$	0.002	0
-10	1.99991	-0.02000	-0.04	0.00200	$+0.00045$
-2	1.76159	-0.00454	-0.008	0.00227	$+0.3130$
-1	1.46212	-0.00274	-0.004	0.00274	$+0.368$
0	1	0	0	0.004	0
0.4011	0.80209	0.00200	0.00160	0.00498	-0.1937
1	0.53788	0.00744	0.004	0.00743	-0.4223
2	0.23840	0.03356	0.008	0.01678	-0.6379
2.128	0.21281	0.04000	0.008512	0.01879	-0.6554
3	0.09485	0.12651	0.012	0.04217	-0.7408
10	0.00009	440.55	0.04	44.052	-0.9091
$+\infty$	0	$+\infty$	$+\infty$	$+\infty$	-1

[a] Z = Coefficient of psychological expansion.

[b] $\psi(Z)[=2/(1+e^Z)]$ = function of desired money balances

[c] $x[=0.002(1+e^Z)Z = \chi Z]$ = rate of increase of global expenditure

[d] $z[=\chi_0 Z = z' = x']$ = psychological rate of expansion

[e] $\chi[=\chi_0/\psi(Z)]$ = rate of forgetfulness

[f] $\epsilon = \dfrac{d\psi}{\psi} \Big/ \dfrac{dx}{x} = -1 \Big/ \left[1 + \dfrac{1}{Z\left(1 - \dfrac{\psi}{2}\right)}\right]$ = elasticity of ψ with respect to x

The maximum value of ϵ is 0.39, and is reached for $Z = -1.3$.

Finally the comparison of the two relations (12) and (14) shows that, if the formulation (14) of the expectation coefficient theory can be considered as a particular case of the formulation (12) of Allais' theory, the opposite proposition does not hold, and relation (12) cannot be derived from relation (14).

In fact, Allais' theory allows no interpretation in the framework of Cagan's expectation coefficient theory.

In the general case—any economic process whatever—the coefficient Z cannot be compared with any of the quantities χ or z, or the product χz. Darby's comment is therefore unfounded.

At the same time, as Z is a zero-dimension coefficient, it is more aptly labelled the "coefficient" rather than the "rate" of psychological expansion. (See Allais 1967, p. 98 and 1969.)

Similarly, Table 1 shows that a rate of expansion x of 4 percent per month results in relative desired cash balances ψ taking on a value of .212, and not .00009 as Darby asserts having assumed the incorrect equality $x = z$.

The economic meaning of the coefficient Z is perfectly clear from the equations [2.2], [2.9], [2.11], and [2.28]. But it may be felt preferable and is probably more appealing intuitively to take relation [2.40].

$$(15)\qquad Z(t) = \int_{-\infty}^{t} x(\tau) \cdot \exp\left[-\int_{\tau}^{t} \chi(u)du\right] d\tau \qquad [2.40]$$

as permitting more direct interpretation. This relation states that Z is the cumulative value of past rates of increase $x(\tau)$ of total outlay, as reduced by forgetfulness, and it is wholly analogous to the equation

$$(16)\qquad V(t) = \int_{t}^{\infty} r(\tau) \cdot \exp\left[-\int_{t}^{\tau} i(u)du\right] d\tau$$

which expresses the present value of a stream of future income $r(\tau)$ as a function of the rate of interest (Allais, 1968b, p. 99).

Relations (15) and (16) considered in conjunction show the profound identity existing between the phenomena of forgetfulness and interest. The past is forgotten in the same way as the future is discounted.

Thus the economic significance of the theory of the demand for money presented in my earlier studies can no doubt be rendered clearer using this new presentation which is mathematically equivalent, but more appealing. The new presentation starts from relation (15) considered as a postulate.

Using this new approach, Z is the central concept and $z=\chi_0 Z$ a derived concept, the meaning of which is easy to see.

It is possible to calculate the constant income v which for a given rate of interest i gives the same discounted present value as equation (16)

$$v = iV \tag{17}$$

Similarly, it is possible to calculate the constant income v_0 which, for a given rate of interest i_0 has a present value of V

$$v_0 = i_0 V \tag{18}$$

Similarly, for Z defined as in (15), it is possible to consider rates χZ and $\chi_0 Z$ for constant rates of forgetfulness χ and χ_0. The interpretation of the products χZ and $\chi_0 Z$ becomes perfectly clear immediately they are considered in analogy to equations (17) and (18).

Darby contends that it is irrational to hold the same amount of cash balances if prices are rising at 5 percent annually with production unchanged, as when production is rising at 5 percent annually with prices unchanged, a point of view shared by Philip Cagan and Jürg Niehans.

There is nothing irrational about this. One's assessment of the economic situation depends on realized profits, and these depend on the evolution of global expenditure. Profits are exactly the same in the two cases envisaged by Darby: in both instances total receipts will exceed total costs by the same percentage.

Consider for example an economy in which the production in period T is the same for all goods and suppose that initially it is in a stationary equilibrium for which the global profit is equal to zero. In such a process total cost is equal to global expenditure. Suppose now that the global expenditure increases from the value $D(\mathrm{t}-\mathrm{T})$ to the value $D(\mathrm{t})$, per unit of time. In the new situation the cost of production is $\mathrm{T}D(\mathrm{t}-\mathrm{T})$ and proceeds are $\mathrm{T}D(\mathrm{t})$. Total profits in the new situation would therefore be equal to the difference between these two quantities. Providing global expenditure rises sufficiently slowly, we have

$$\mathrm{T}[D(\mathrm{t}) - D(\mathrm{t} - \mathrm{T})] \sim \mathrm{T}^2 \frac{dD}{d\mathrm{t}} \tag{19}$$

and it can be seen that in this particular case, profits depend only on $dD/d\mathrm{t}$, whatever the relative contributions of the price rise and production rise components to

$$x = \frac{1}{D}\frac{dD}{d\mathrm{t}} = \frac{1}{P}\frac{dP}{d\mathrm{t}} + \frac{1}{Q}\frac{dQ}{d\mathrm{t}} \tag{20}$$

At all events, the Allais formulation is unduly and incorrectly restricted when in relation (15), x denotes the rate of increase of prices rather than the rate of increase of global expenditure. It follows that the findings *are what one would expect if* economic agents as a body only took the growth rate of global expenditure into account.

In fact any realistic theory must bow to the factual record, rather than to a priori views however appealing to those who formulate them (on all these points see my reply to comments by Cagan and Niehans, especially § A5).

Allais replaced x the rate of increase of global expenditure by p, the rate of price increase, for hyperinflations because there are no monthly national income data and because there is no drawback to the assumption that production is constant in hyperinflations.

There is no reason for the rate of increase of production to be taken into consideration for the purpose of selecting the values of the fitting parameters ϕ_0 and k. This is obvious for k, since it is easily shown that the influence of this constant declines progressively with the passage of time. As far as ϕ_0 is con-

cerned, it is only used in order to fix the level of the ratio M/R of the money supply to national income, and there seems no reason why this ratio should be dependent on the rate of growth of real income.

The elasticity ϵ of ψ with respect to x *in a state of dynamic equilibrium* is easily calculated. From equation (3)

$$\frac{dZ}{Z} - \frac{d\psi}{\psi} = \frac{dx}{x} \tag{21}$$

and, using (2.42) of [5], it is possible to deduce

$$\begin{aligned} \epsilon &= \frac{d\psi}{\psi} \bigg/ \frac{dx}{x} \\ &= -1 \bigg/ \left[1 + \frac{1}{Z\left(1 - \frac{\psi}{2}\right)}\right] \end{aligned} \tag{22}$$

This elasticity ϵ is not constant (see Table 1), and it can be seen that it does not even remain of the order of -0.3, as Darby asserts, but varies between $+0.39$ and -1 when x varies between $-\infty$ and $+\infty$. For $x=0.75$ percent per month (8.9 percent per annum), ϵ is equal to -0.42 (Table 1).

In itself, and assuming that it is justified—which it is not—Darby's argument that an elasticity of ψ with respect to x of the order of -0.4 is plausible, is hardly of any scientific value. A formulation is not valid as a function of intuitive considerations which can differ from one author to the next, but only in the light of the exactness with which it is borne out through confrontation with the observed data. The basic argument in favor of the theory of the demand for money that I have proposed lies here, nowhere else.

Darby seems to hint that it might have been better to determine the parameters χ_0 and α for each particular case rather than to ascribe values for them through postulates.

In the first instance, I would point out that the fittings in which these parameters were considered as variable were in fact carried out. For the fifteen series considered, the averages of the values found for χ_0, α and b turned out to be approximately equal to the values 0.004, 1, and 1, indicated by the postulates.

But secondly, surely a theory is all the more noteworthy when it succeeds in bringing the greatest possible number of facts within the sweep of a formulation common to all the cases studied, and most of whose parameters are the same in each of these cases. (See Allais 1965, pp. 141–45 and 157–58; and pp. 1149–50 and 1156 of my *Review* article.)

Finally, the verification of the theory presented using constant values of the parameters α, b, and χ_0 shows that the function of the demand for money is *stable from one country to another and from one period to another*. In the light of the issues under debate in these last twenty years, this is a result of considerable importance.

For all these reasons Darby's point of view does not appear as justified.

Like many English-language writers, Darby seeks at all costs to reduce the theory presented to the concepts and theories which have found acceptance to date, at least in some circles, even though these are not well verified empirically (see for example Niehans' and Cagan's comments on my paper).

This is a mistake, and Allais' theory can only be properly understood if the reader is prepared to shed the concept of anticipation, and base his reasoning on the consideration of new concepts, namely the rate of forgetfulness, psychological time, the hereditary influence of the past, the coefficient of psychological expansion Z, etc.

In point of fact, the rate Z defined by the differential equation

$$\frac{dZ}{dt} + \frac{\chi_0}{\psi(Z)} Z = x(t) \tag{23}$$

does not define expectations at all. It represents the hereditary influence of past value of the rate x as defined in relation (15) above.

The consideration of the influence of the hereditary effects of past events on present decisions is more fruitful than an attempt to explain present decisious in terms of expectations about the future. Expectations about the future can only be explained by the consideration of the hereditary effects of past

events, and once this is accepted, *they constitute an unnecessary intermediary stage in the explanation of present decisions.*

It is quite doubtful that my theory can be adequately represented or interpreted in the jargon currently used by some theorists of monetary dynamics. The words used to indicate concepts are unfortunately not interchangeable and certain expressions can suggest false ideas. For example, it is possible to interprete expectations as a function of the hereditary effects of past events, but to me it does not seem that a useful analysis of hereditary effects can be obtained by reducing their interpretation to that of the analysis of expectations. So to try to interpret with the current jargon the theory I have proposed can only give rise to inaccurate ideas.

Darby proposes that the relation

$$\frac{dz}{dt} = \chi_0\left[x - \frac{z}{\varphi(z)}\right] \tag{24}$$

equivalent to relation (23) be replaced by

$$\frac{dz}{dt} = \frac{\chi_0}{\varphi}[\varphi(z)(x - z) + [1 - \varphi(z)](\bar{x} - z)][5] \tag{25}$$

where $\bar{x}$ represents the "normal value" of x, the two relations being equivalent for $\bar{x}=0$.

This relation is obviously more general than (24) and will no doubt produce superior fits of the data since it involves one more degree of freedom. However, the values of the coefficients of correlation derived from Allais' theory, relation (24), are so high that it is doubtful whether there would be any sensible gain of explanatory power.[6] The best theory is the simplest one, the one with the fewest explanatory parameters for the same explanatory power.

Further, it is difficult to see why one should take account of a weighted average of $(x-z)$ and $(\bar{x}-z)$ whose weights happen to be precisely equal to φ and $1-\varphi$.

The only reason we can see for the alternative formulation (25) is Darby's a priori desire to have the new theory brought onto the same ground as the theory of coefficients of expectations. This seems difficult to rationalize—except possibly though the tendency of the human mind to accept certain propositions as proven verities on the simple grounds that everybody else accepts them as such without proof.

In reality, the coefficient of psychological expansion Z lends itself to the very simple interpretation given by relation (15) which is analogous to the formulation (16). It brings out the fundamental circumstance that there is a formal analogy, and indeed a functional analogy, between the rate of forgetfulness and the rate of interest. One serves to calculate the residual cumulative influence of the past, the other the discounted cumulative influence of the future.

It is also worth mentioning that the analogy between relations (15) and (16) reinforces the postulate according to which the rate of forgetfulness is equal to the rate of interest, i.e.,

$$\chi = i \tag{26} \qquad [2.49]$$

This relation, using (2.39) and (2.51) of [5], leads to

$$\frac{i}{i_0} = \frac{1 + e^Z}{2} \tag{27}$$

It can be seen that this relationship is remarkably well borne out by an analysis of the yield of Consols for the period 1815 to 1913 (on all these points see Allais (1968), (1968b, p. 128), (1969 Appendix B, § III) and (1969b § A6, A11, B3 and B14).

In any case, and as I intend to show in a forthcoming paper, the introduction in the monetary theory of a value $\bar{x}$ of x, which can be considered as normal, can be extremely useful. But in my opinion this new concept does play a role, very different of that suggested by Friedman's in Darby's relation (25). The rate $\bar{x}$ does represent what indeed we may call the rate of expected increase of global expenditure when the coefficient Z (and the rates $\chi_0 Z$ and χZ which correspond

[5] I am using $\bar{x}$ in place of Darby's k to avoid possible confusion with the parameter k of relation [3.18] of my *Review* article.

[6] See my *Review* article, pp. 1142–42.

to it) is an index representing the collective appreciation by the economic agents of the conjunctural situation.

Darby comments that it seems implausible to have simultaneously

$$(28)\qquad \frac{1}{\phi_0}\frac{M_D}{R} = \psi(Z)$$

$$(29)\qquad \frac{i}{i_0} = \psi(Z)$$

What is implausible about it? The theory proposed is so well verified by the data that *the record reads as it would* if both relations were valid. Further, they are both the consequence of very natural-seeming postulates.

In any case the empirical verification of (28) is excellent (see my *Review* article pp. 1141–42), and so far as relation (29) is concerned, it is also remarkably verified.

From (28) and (29) we have

$$(30)\qquad \frac{iM_D}{R} = \frac{\phi_0}{i_0}$$

which means that the elasticity of M_D/R with respect to the psychological rate of interest is -1. This result is quite appealing.

Darby considers that an advantageous feature of Cagan's demand function is that it has no maximum. In point of fact, the opposite holds. We have

$$(31)\qquad \frac{M_D}{R} = \frac{M_D}{C}\frac{C}{R}$$

where M_D denotes desired cash balances, R national income, and C national wealth.

M_D/C can be interpreted as the share of their assets which operators prefer to hold in liquid form. The upper bound of this ratio is unity. The ratio C/R of national capital to national income is also bounded from above.[7] Hence M_D/R has an upper limit.[8]

Instead of

$$(32)^9\qquad \varphi(z) = \frac{1+b}{1+be^{az}}$$

Darby proposes the alternative formulation

$$(33)\qquad \varphi(z) = \frac{1+b}{1+be^{\mu|z|}}$$

in which the absolute value $|z|$ of z replaces its algebraic value. It is hard to see how this suggestion could be justified. It implies that in a hyperdeflation M_D/R would tend to zero. All the available evidence contradicts this assumption. Be this as it may, it would be necessary to show that (32) leads to better fittings than the present formulation (31). This proof has yet to be supplied.

Thus Darby's comments on the points which he raised rest on wrong interpretations of Allais' theory and are not justified, or else are grounded on a priori assumptions whose empirical justification has yet to be given.

Some of the Darby's arguments may appear, at least from his viewpoint, very appealing, but, a priori, an intuition is nothing more than an intuition. It is possible to decide between two different intuitions, notably Darby's and mine, only by confronting them with reality. This is the decisive argument, because a theory is of value to the extent that it is verified by the available empirical data, and only to that extent. The most contradictory viewpoints have had support in economics. Neither intuition, reasoning, or even solidly established and consecrated usage can distinguish their relative merits. Experience and experience alone can do this. It is no doubt difficult to discard an old viewpoint whose only real justification may be that we have grown accustomed to it and that it has been widely adopted. But this should be done nonetheless if a new approach, less attractive because we are not yet used to it, can represent reality better. As far as the new theory is concerned, the only fact which has been solidly established is that in every case studied to now, whether normal, inflationary, deflationary or hyperinflationary situations have been examined, the observed data can be represented with remarkable fidelity *by one and the same formulation*. In each case this formulation depends on only

[7] See Allais 1962 Table 1.

[8] See Allais 1956 §51 A and §50A, p. 231.

[9] Relation equivalent to (2.35) of my *Review* article.

two arbitrary parameters, which represent the constants of integration. Naturally this accuracy of representation does not prove the exactitude of the postulates on which the new formulation of the demand for money is based. No theory, no matter how impressive its empirical verification, can do this. We can never pretend to fathom the depths of a phenomenon. All we can do is to propose models to explain what is happening, and of two models, the best is the one which represents reality at the same time better and more simply. The results obtained show only, that in the present state of things and so long as no alternative theory giving better results has been proposed, it may be considered that *the course of events is as it would be if* the proposed formulation, and the postulates on which it is based correctly represent the nature of the phenomena observed. If I may paraphrase Darby's conclusion, I shall myself conclude that in summary the particular form of Darby's criticisms is resting on wrong interpretations of Allais' theory or on highly questionable a priori hypotheses, but that a reformulation of his criticisms could be very promising.

Appendix

New Formulation of Allais' Theory

Because of the fundamental analogy between forgetfulness and interest, and in order to mark its importance, I have adopted since November 1967 a new formulation of the postulates of my theory of the demand for money.[10] This new formulation is *mathematically equivalent* to the preceding one, but is certainly simpler. It throws greater light on the economic significance of my theory and clarifies the preceding discussion.

The new formulation of monetary demand which I have proposed can be based on the following considerations (see Allais, 1968, and 1968b pp. 97–108):

a) At any moment t the past is taken into account in the same way as the future. The influence of the past on the present is reduced by forgetfulness, just as the influence of the future on the present is discounted. At any moment, it is possible to define for the whole of the economy rate of instantaneous forgetfulness $\chi(t)$ which plays the same role vis-à-vis memory as that played by the interest rate $i(t)$ in discounting the future.
b) The overall appraisal of the economic situation by the collectivity can be represented by a coefficient Z, called the "coefficient of psychological expansion," which represents the cumulative value of the influences (reduced by forgetfulness) of the rates x of growth of aggregate expenditure observed in the past. Z is defined by equation (15) above.
c) A psychological time t' can be defined to correspond to physical time t under the condition that the corresponding rate of forgetfulness χ' per unit of psychological time is a constant equal to the value χ_0 of the rate of forgetfulness for $Z=0$ (relation [2.2]).

   When psychological time is considered, the velocity of circulation of total desired cash balances, if measured using this time as frame of reference, is assumed to be constant.

With this new presentation the postulate [2.11] of my *Review* article becomes useless and the relation [2. 12] *disappears*. It is simply replaced by equation (1) above which remains unchanged.

If we put

$$\text{(A1)}\quad Z' = \int_{-\infty}^{t'} x'(\tau') \exp\left[-\int_{\tau'}^{t'} \chi' du'\right] d\tau'$$

we see immediately that according to [2.2] and [2.8] of we have

$$\text{(A2)}\qquad Z = Z'$$

Thus Z is invariant, i.e., independent of the time scale used for reference. Relation (A 2) replaces relations (2.11) of, but relation (A-2) is no longer a postulate.

In addition, I introduced a new concept corresponding to the product χZ, denoting this product χZ by z and the product $\chi_0 Z$ by z_0.[11]

[10] This new presentation will appear in a forthcoming paper.

[11] To avoid any misinterpretation, the use of these new notations is strictly limited to this Appendix.

The correspondence between the present and earlier notations is as follows:

Earlier notation[12]	*Present notation*
$z' = z = \chi_0 Z$	$z' = z_0 = \chi_0 Z$
χZ	$z = \chi Z$

With the introduction of the new concepts Z' and $z=\chi Z$ and *with the new notation* we have the following correspondence between the physical time scale and the psychological time scale.

Time Scale

	Physical		Psychological	
(A3)	ψ	=	ψ'	
(A4)	Z	=	Z'	
(A5)	χ		$\chi' = \chi_0$	[2.22]
(A6)	$z = \chi Z$		$z' = \chi' Z'$	
			(or $z_0 = \chi_0 Z$)	
(A7)	$x dt$	=	$x' dt'$	[2.8]
(A8)	χdt	=	$\chi' dt'$	[2.2]
(A9)	$z dt$	=	$z' dt'$	

Relation (2) above naturally remains unchanged.

As far as the new concept $z=\chi Z$ is concerned, we have according to (12)

$$\frac{dz}{dt} = \chi(x - z) + \frac{z}{\chi}\frac{d\chi}{dt} \tag{A10}$$

The nature of (A 10) is different from that of relation (2) above, but, as we have seen,[13] *in a dynamic equilibrium process* where χ remains constant, this relation becomes

$$\frac{dz}{dt} = \chi(x - z) \tag{A11}$$

Thus the parameter z appears to play a role similar to the Cagan's rate of expectation. *But relation (A 11) is valid only for a dynamic equilibrium process, when relation* (2) *above holds in any case.*

Given the parallelism between forgetfulness and interest, the new notations have the advantage of being more suggestive. In addition they are liable to help to avoid possible erroneous interpretations, such as that of Darby.

According to an earlier definition $v_0 = i_0 V$ is equal to the constant future income whose global value, actualized with a constant rate of interest i_0, would be equal to the actualized value V of the future incomes $r(t+\theta)$ for rates of interest equal to $i(t+\theta)$ (relation 16 above).

Likewise $v = iV$ corresponds to the rate i of the considered instant t as v_0 corresponds to the rate i_0.

We have identically

$$V(t) = \int_t^{+\infty} i(t)V(t) \cdot \exp[-i(t)(\tau - t)]d\tau \tag{A12}$$

and an analogous equation for the product $i_0 V$.

In the same way $z_0 = \chi_0 Z$ is equal to the constant past rate of increase of the global expenditure whose present influence with a constant rate of forgetfulness χ_0, would be equal to the global present influence Z of the previous rates x of increase of the global expenditure, the previous rates of forgetfulness having the values $\chi(u)$, relation (15) above.

Likewise $z = \chi Z$ corresponds to the rate of forgetfulness χ of the instant t considered as z_0 corresponds to the rate χ_0.

We have identically

$$Z(t) = \int_{-\infty}^{t} \chi(t)Z(t) \cdot \exp[-\chi(t)(t - \tau)]d\tau \tag{A13}$$

and an analogous equation for the product $z = \chi Z$.[14]

REFERENCES

M. Allais, "Explanation of Economic Cycles by a Non-Linear Monetary Model With

[12] In my *Review* article no notation corresponded to the product χz, whose consideration appears to be quite interesting.

[13] See relation (12) above.

[14] On these points see Allais, (1968) and (1968b) pp. 98–99.

Lagged Relations," (Explication des Cycles Economiques par un Modèle non Linéaire à Régulation Retardée), *Metroeconomica*, Apr. 1956, *8*, 4–83; also in *Les Modèles Dynamiques en Econométrie* (Dynamic Models in Econometrics), National Center of Scientific Research (C.N.R.S.), Collection of Proceedings of International Colloquia, 1956, Vol. 62, pp. 169–308.

———, "The Influence of the Capital-Output Ratio on Real National Income," Walras-Bowley Lecture, *Econometrica*, Oct. 1962, *30*, also in K. J. Arrow and T. Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969, pp. 682–714.

———, *La Reformulation de la Theorie Quantitative de la Monnaie*, Editions SEDEIS, Paris 1965.

———, "A Restatement of the Quantity Theory of Money," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1123–57.

———, "Forgetfulness and Interest," (Oubli et Intérêt), lectures delivered at the Graduate Institute of International Studies, Geneva, Dec. 14-15, 1967; and at Nuffield College, Oxford University, Mar. 8–11, 1968, (to be published).

——— (1968b), "Money and Growth," (Monnaie et Developpement), Vol. II, Ecole Nationale Superieure des Mines de Paris, September 1968.

———, "Growth and Inflation," *J. Money, Credit, Banking*, Aug. 1969, *1*, 355–426.

——— (1969b), "Growth and Inflation, A Reply to the Observation of the Discussants," *J. Money, Credit, Banking*, Aug. 1969, 441–62.

P. Cagan, "Allais' Monetary Theory; Interpretation and Comment," *J. Money, Credit, Banking*, Aug. 1969, *1*, 427–32.

J. Niehans, "Growth and Inflation, A Comment," *J. Money, Credit, Banking*, Aug. 1969, 433–38.

Third-Degree Stochastic Dominance

By G. A. WHITMORE*

In the March 1969 issue of this *Review*, Josef Hadar and William Russell reported the following two rules for ordering a pair of uncertain prospects $F(x)$ and $G(x)$. They called the two conditions that formed the basis of these rules, first- and second-degree stochastic dominance, respectively.

Rule 1: The prospect $F(x)$ is not preferred to the prospect $G(x)$ if, and only if, $F(x)-G(x)\geq 0$ for all $x\in[a, b]$.

Rule 2: The prospect $F(x)$ is not preferred to the prospect $G(x)$ if, and only if, $\int_a^x (F(y)-G(y))dy\geq 0$ for all $x\in[a, b]$.

Here $F(x)$ and $G(x)$ are less-than cumulative probability distributions where x is a continuous or discrete random variable representing the outcome of a prospect. The closed interval $[a, b]$ is the sample space of both prospects. The integral shown in *Rule* 2 and those shown throughout the paper are Stieltjes integrals. Recall that the Stieltjes integral $\int_a^b f(x)dg(x)$ exists if one of the functions f and g is continuous and the other has finite variation in $[a, b]$.

Let D_1, D_2, and D_3 be three sets of utility functions $\phi(x)$. D_1 is the set containing all utility functions with $\phi(x)$ and $\phi_1(x)$ continuous, and $\phi_1(x)>0$ for all $x\in[a, b]$. D_2 is the set with $\phi(x)$, $\phi_1(x)$, $\phi_2(x)$ continuous, and $\phi_1(x)>0$, $\phi_2(x)\leq 0$ for all $x\in[a, b]$. D_3 is the set with $\phi(x)$, $\phi_1(x)$, $\phi_2(x)$, $\phi_3(x)$ continuous, and $\phi_1(x)>0$, $\phi_2(x)\leq 0$, $\phi_3(x)\geq 0$ for all $x\in[a, b]$. Here $\phi_i(x)$ denotes the ith derivative of $\phi(x)$. Hadar and Russell proved that *Rule* 1 is valid for all $\phi\in D_1$ and *Rule* 2 is valid for all $\phi\in D_2$.

The authors point out that the set of probability distributions that can be ordered by means of second-degree stochastic dominance is, in general, larger than that which can be ordered by means of first-degree stochastic dominance. Note that in *Rule* 2, they assume that $\phi(x)$ is not only an increasing function of x but also exhibits weak global risk aversion, a condition guaranteed by requiring the second derivative of $\phi(x)$ to be nonpositive.

In this paper, a condition which will be called third-degree stochastic dominance is considered. It is based on the following assumption about the form of the utility function $\phi(x)$. From a normative point of view, one expects the risk premium associated with an uncertain prospect to become smaller the greater is the individual's wealth. The plausibility and implications of this assumption have been explored by John Pratt, as well as others. The risk premium of an uncertain prospect is that amount by which the certainty equivalent of the prospect differs from its expected value. In mathematical terms, given the prospect $F(x)$ with expected value μ, the corresponding risk premium π is obtained by solving the following equation.

$$(1) \qquad \phi(\mu-\pi) = \int_a^b \phi(x)dF(x)$$

The foregoing assumption states that π is a nonincreasing function of μ. Pratt shows that this condition is equivalent to the condition that $r(x)=-\phi_2(x)/\phi_1(x)$ is a nonincreasing function of x, that is, $r'(x)\leq 0$. Since:

$$(2) \quad r'(x) = (-\phi_3(x)\phi_1(x) + \phi_2(x)^2)/\phi_1(x)^2$$

the requirement that $r'(x)\leq 0$ is equivalent to $-\phi_3(x)\phi_1(x)+\phi_2(x)^2\leq 0$. If it is assumed that $\phi_1(x)>0$, then the requirement that $r'(x)\leq 0$ implies $\phi_3(x)\geq 0$. Note that it is necessary for $\phi_2(x)\leq 0$ in order for $r(x)$ to be nonnegative. Also, note that since $\phi_3(x)\geq 0$ is only a necessary condition for $r'(x)\leq 0$, it follows that D_3 contains utility functions where both $r'(x)\leq 0$ and $r'(x)>0$.

Having made this observation, it is possible to prove the following theorem.

Theorem: The prospect $F(x)$ is not preferred to the prospect $G(x)$ for all utility functions in D_3 if, and only if:

* Associate professor, McGill University, Montreal.

$$\text{(3a)} \qquad \int_a^x \int_a^y (F(z) - G(z))dzdy \geq 0 \quad \text{for all } x \in [a, b]$$

and

$$\text{(3b)} \qquad \int_a^b (F(y) - G(y))dy \geq 0$$

A proof of this theorem, which parallels those used by Hadar and Russell, is given in the Appendix. The conditions (3a) and (3b) define third-degree stochastic dominance. The theorem represents a third rule for ordering uncertain prospects. Second-degree stochastic dominance implies third-degree stochastic dominance. Thus, the set of distributions which can be ordered by means of third-degree stochastic dominance is, in general, larger than that which can be ordered by means of second-degree stochastic dominance.[1]

The theorem for third-degree stochastic dominance can be stated in the following equivalent form.

Theorem: The prospect $F(x)$ is not preferred to the prospect $G(x)$ for all utility functions in D_3 if, and only if:

$$\begin{aligned} &x^2[F(x) - G(x)]/2 \\ &\quad - x\left[\int_a^x ydF(y) - \int_a^x ydG(y)\right] \\ \text{(4a)} \quad &\quad + \left[\int_a^x y^2/2dF(y)\right. \\ &\quad \left. - \int_a^x y^2/2dG(y)\right] \geq 0 \\ &\quad \text{for all } x \in [a, b] \end{aligned}$$

and

$$\text{(4b)} \qquad \int_a^b ydG(y) - \int_a^b ydF(y) \geq 0$$

Since means and variances frequently are employed in economic studies to make comparisons between uncertain prospects,[2] it is interesting to note that third-degree stochastic dominance places some restrictions on the associated means and variances. In particular, if μ_F, μ_G, σ_F^2, and σ_G^2 denote the means and variances of $F(x)$ and $G(x)$, respectively, then by setting $x=b$ in (4a), the following two conditions are obtained.

$$\text{(5a)} \qquad (\sigma_F^2 - \sigma_G^2) + (\mu_G - \mu_F)(2b - \mu_G - \mu_F) \geq 0$$

$$\text{(5b)} \qquad \mu_G - \mu_F \geq 0$$

The expressions in (5a) and (5b) are necessary (but not sufficient) conditions on the means and variances of $F(x)$ and $G(x)$ for $G(x)$ to be at least as preferred as $F(x)$ for all utility functions in D_3.

Although the ordering strength of third-degree stochastic dominance exceeds that of second-degree stochastic dominance, there remain many practical situations where the preference between two uncertain prospects cannot be established by this condition. For instance, the following situation frequently arises in practice. A prospect $G(x)$ appears attractive relative to a second prospect $F(x)$ except that one (or several) outcomes of $G(x)$ fall below all the possible outcomes of $F(x)$. Regardless of how attractive $G(x)$ is and how small the probabilities of the adverse outcomes of $G(x)$ are (provided they are not zero) the rule for third-degree stochastic dominance will not indicate that $G(x)$ is at least as preferred as $F(x)$.

Nevertheless, in spite of the fact that third-degree stochastic dominance will not provide an ordering in many practical situations, the concept should have considerable value in theoretical investigation.

Appendix

The proof of the theorem proceeds in two stages: First, to verify the implication that if $\int_a^x\int_a^y (F(z)-G(z))dzdy \geq 0$ for all $x \in [a, b]$ and $\int_a^b (F(y)-G(y))dy \geq 0$ then prospect $F(x)$ is not preferred to prospect $G(x)$ for all $\phi \in D_3$. Second, to verify the converse of the foregoing implication, let $H_n(x) = \int_a^x H_{n-1}(y)dy$

[1] If the set of utility functions is restricted to only those for which $r'(x) \leq 0$, then by employing the three conditions $\phi_1(x) > 0$, $\phi_2(x) \leq 0$, and $-\phi_3(x)\phi_1(x) + \phi_2(x)^2 \leq 0$, a stronger ordering rule than third-degree stochastic dominance can be obtained. The author has not attempted to construct this stronger rule.

[2] For instance, see H. M. Markowitz and James Tobin.

for $n>1$ where $H_1(x)=F(x)-G(x)$. Carrying out several integrations by parts, the following relationship is obtained.

$$
\begin{aligned}
&\int_a^b \phi(x)dH_1(x)\\
\text{(A-1)}\qquad &= \left| \phi(x)H_1(x) \right|_a^b - \left| \phi_1(x)H_2(x) \right|_a^b\\
&\quad + \left| \phi_2(x)H_3(x) \right|_a^b\\
&\quad - \int_a^b \phi_3(x)dH_4(x)
\end{aligned}
$$

Since $H_n(a)=0$ for $n\geq 1$ and $H_1(b)=0$, (A-1) may be written:

$$
\begin{aligned}
&\int_a^b \phi(x)dH_1(x)\\
\text{(A-2)}\qquad &= -\phi_1(b)H_2(b)+\phi_2(b)H_3(b)\\
&\quad - \int_a^b \phi_3(x)dH_4(x)
\end{aligned}
$$

Since $\phi_1(x)>0$, $\phi_2(x)\leq 0$, and $\phi_3(x)\geq 0$ for all $x\in[a, b]$, it follows that $(A-2)$ is nonpositive if $H_2(b)\geq 0$ and $H_3(x)\geq 0$ for all $x\in[a, b]$. Hence, the first stage of the proof is complete.

To prove the converse, a proof by contradiction is employed. Suppose that for some constant c, $a<c\leq b$, $H_3(x)\geq 0$ for $c<x\leq b$ and $H_3(x)<0$ for $a\leq x\leq c$ and/or $H_2(b)<0$. In this case, it can be shown that there exists $\phi\in D_3$ for which $F(x)$ is preferred to $G(x)$. For instance, consider the following utility function.

$$
\text{(A-3)}\quad \phi^*(x)=\begin{cases} P(x)+Q(x) & a\leq x\leq c\\ P(x) & c<x\leq b\end{cases}
$$

$$
\begin{aligned}
\text{where}\quad P(x) &= -A(c-a)^2x^2/4\\
&\quad + B(c-a)^2bx/2 + C\\
\text{and}\quad Q(x) &= -D(c-x)^4/24
\end{aligned}
$$

The parameters A, B, C, and D are arbitrary constants chosen freely subject to the conditions $B>A\geq 0$ and $D\geq 0$. An inspection of $\phi^*(x)$ indicates that $\phi^*(x)\in D_3$. Three cases must be considered here. These are listed below.

$$
\begin{aligned}
\text{Case (a)}:\ & H_3(x)\geq 0 \quad \text{for } c<x\leq b,\\
& H_3(x)<0 \quad \text{for } a\leq x\leq c,\\
& H_2(b)\geq 0\\
\text{(A-4)}\qquad \text{Case (b)}:\ & H_3(x)\geq 0\\
&\qquad\qquad \text{for all } x\in[a, b],\\
& H_2(b)<0\\
\text{Case (c)}:\ & H_3(x)\geq 0 \quad \text{for } c<x\leq b,\\
& H_3(x)<0 \quad \text{for } a\leq x\leq c,\\
& H_2(b)<0
\end{aligned}
$$

For $\phi^*(x)$, after some rearrangement, $(A-2)$ may be written:

$$
\begin{aligned}
&\int_a^b \phi^*(x)dH_1(x)\\
\text{(A-5)}\qquad &= ((A-B)(c-a)^2b/2)H_2(b)\\
&\quad - (A(c-a)^2/2)H_3(b)\\
&\quad - \int_a^c D(c-x)dH_4(x)
\end{aligned}
$$

In Case (a), by holding A and B constant and making D sufficiently large, $(A-5)$ becomes positive. This is a contradiction of the requirement that $F(x)$ be not preferred to $G(x)$. Similarly, in Cases (b) and (c), by holding A and D constant and making B sufficiently large, $(A-5)$ becomes positive. Again, this is a contradiction. Hence, the second stage of the proof is complete.

The above proof has assumed implicitly that $b>0$. This is the situation that is encountered most frequently. However, the theorem also holds for $b<0$ and $b=0$, although the argument of the proof must be modified slightly in these two cases.

REFERENCES

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *49*, 25–34.

H. M. Markowitz, *Portfolio Selection—Efficient Diversification of Investments*. New York 1959.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan.-Apr. 1964, *32*, 122–36.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. Econ. Stud.*, Feb. 1958, *26*, 65–86.

Does Futures Trading Reduce Price Fluctuations in the Cash Markets?

By Mark J. Powers*

One of the recurring arguments made against futures markets is that, by encouraging or facilitating speculation, they give rise to price instability. This argument, in various versions, has been made throughout past Congressional hearings on onion and potato futures. The theoretical literature available on the subject of futures trading and price stability is rather scanty and inconclusive. Most of the evidence has been gathered on onion and potato prices. It suggests that: a) the seasonal price range is lower with a futures market because of speculative support at harvest time; b) sharp adjustments at the end of a marketing season are diminished under futures trading because they have been better anticipated; and c) year-to-year price fluctuations are reduced under futures trading because of the existence of the futures market as a reliable guide to production planning. See Roger Gray, and Holbrook Working (1958, 1960, 1963).

These conclusions are most valid for seasonally produced storable commodities. They probably do not hold for other commodities, particularly those that are continuously produced and semi- or non-storable. A more general approach to the study of the impact of futures trading on cash prices is needed.

The statistical evidence assembled in support of the above three conclusions has always dealt with the total seasonal variation in cash prices. This paper concerns itself with an analysis of the impact of futures trading on the fluctuations of the separate elements of a price series. It focuses mainly on the effect futures trading has on the random element.

* Associate professor of economics, South Dakota State University. I am indebted to Roger Gray of the Stanford Food Research Institute, and A. C. Johnson of the department of agricultural economics, University of Wisconsin, for comments on an earlier draft of this paper. Errors in the paper are entirely the responsibility of the author.

Working (1934 p. 11–24) and Arnold Larson (p. 313–24) have provided evidence that an anticipatory price series may be thought of as describing a random walk. Paul Cootner and others in analyses of stock market prices have done additional work on the hypothesis. Hendrik Houthakker (p. 164–72) questions the applicability of the random-walk hypothesis to futures prices, particularly in the short run. He argues that futures price series are composed of both a random element and a systematic component. None of these studies considers the application of the random-walk hypothesis to the cash price series.

If one were to accept Working's and Larson's results as being applicable to a cash price series, then a measure of the total variation in the series would also be a measure of the random element in the series. On the other hand, if their evidence on futures price series is not relevant to cash price series, then a measure of the total variation in the cash series would be a measure of more than just the random element. Since cash price series are not entirely anticipatory and no evidence is available that they do or do not describe a random walk, it is assumed that the latter is correct and that the cash price series considered here are composed of both a systematic and a random component.

The first part, the systematic component of the time-series, is the part of the price series that is representative of fundamental economic conditions and varies with those conditions. Fluctuations in this component are not only desirable but necessary in a free market economy where prices allocate resources and distribute outputs. The second part of the price series, the random component, cannot be explained by underlying economic forces. Random fluctuations are undesirable in a price series because they act only to distort the price message and, thus,

really represent a "noise" factor in the price system.

To illustrate, consider:

$$P_t = S_t + E_t \tag{1}$$

The variance of the series is

$$V(P_t) = V(S_t) + V(E_t) \tag{2}$$

and the covariance is

$$\mathrm{Cov}\ (S_t E_t) = 0 \tag{3}$$

where

P_t = time-series of prices
S_t = the systematic component associated with fundamental economic conditions
E_t = the error or random component which represents noise and disturbance in the price system.

One would expect that $V(S_t) \neq 0$, implying that changes in economic conditions result in price changes, and it would be desirable that $V(E_t) = 0$, i.e., that there be no random fluctuations.

Given the above, then the real concern with the pricing system is to improve its efficiency as a communications and allocative device by reducing the noise factor, the random elements in the prices, while maintaining the responsiveness of the systematic component of fundamental economic conditions.

Since a price series is composed of systematic and random parts, the question arises: What would happen to these two components if a viable futures market were introduced into the pricing system? Consider $V(S_t)$ first.

One possibility is that the introduction of futures trading would have no effect on $V(S_t)$. A second and more likely result might be that $V(S_t)$ would be reduced. Gray (p. 273–75) and Working (1958 and in Congressional hearings) support this contention. This might occur because of a close relationship between expected (futures) prices and economic fundamentals.

In general, however, there have been few studies of the impact of futures trading on $V(S_t)$, since most studies do not distinguish between $V(S)$ and $V(E)$. Important questions with respect to $V(S_t)$ remain unanswered, suggesting further research is needed in this area.

The main focus of this paper, however, is $V(E_t)$, the second element of variance in the price series. What happens to $V(E_t)$ when a futures market is injected into the marketing system for a commodity?

To seek an answer to this question an analysis of two commodities, pork bellies and live cattle, was undertaken. For purposes of conducting the study it was hypothesized that $V(E_t)$ would be lower during time periods with futures trading than during time periods without futures trading.

To analyze the data and test the hypothesis it was desirable to use a technique which would isolate and estimate the random element in a variable which is changing over time. The technique ultimately selected was the *Variate Difference Method*, developed by Gerhard Tintner.

I. The Variate Difference Method

Although several different techniques night be used for determining the random variations in a time-series, the variate difference method fits our purpose best, mainly because it is a statistical method that does not require the specification of a rigid model and it isolates and estimates the random element without affecting the systematic component.

The variate difference method starts from the assumption that an economic time-series consists of two additive parts. The first is the mathematical expectation or systematic component of the time-series. The second is the random or unpredictable component. The assumption is that these two parts are connected by addition but are not correlated. It is further assumed that the random element is not autocorrelated and has a mean of zero; that the random element is normally distributed; and that the systematic component is a "smooth" function of time.

The steps involved in the analysis are essentially three. First, the random element is isolated in the time-series. This is accomplished by finite differencing. Successive finite differencing of a series will eliminate or at

least reduce to any desired degree the systematic component without changing the random element at all. The random component cannot be reduced by finite differencing because it is not ordered in time.

Second, the variance of the random element is calculated. The calculation of the variance for the original series is as follows:

$$V_0 = \frac{\sum_{i=1}^{N} (W_1 - \overline{W})^2}{N - 1}$$

where $\overline{W}$ = the mean of the original series.

The estimate of the variances of the higher differences is as follows:

$$V_k = \frac{\sum_{i=1}^{N-K} (\Delta^{(k)} W_i)^2}{(N - K)_{2k}C_k}$$

where ${}_{2k}C_k$ is the binomial coefficient of the K^{th} difference equal to the number of combinations of $2k$ things taken k things at a time.

When a finite difference of the order K_o is found such that the variance of the K_o^{th} difference is equal to the variance of the $(K_o+1)^{\text{th}}$ difference and equal to that of the $(K_o+2)^{\text{th}}$ difference, and equal to the $(K_o+3)^{\text{th}}$ difference etc., it is reasonable to assume that the mathematical expectation has been eliminated to a reasonable degree by taking K_o differences and that the remaining variance is attributable to the random element. The variances are considered to be approximately equal, from a probability standpoint, when the differences between the variances for successive finite differences are less than three times the standard error.

To determine whether or not there is a statistically significant difference in the random variance for price series in different time periods, a standard-error-difference formula for testing the difference between two variances was used.[1]

II. *The Analysis*

Weekly cash prices for pork bellies and live beef were collected for eight years, four years preceding the start of futures trading and four afterwards.[2] The four-year periods considered for pork bellies were 1958 through 1961, and 1962 through 1965. For beef, the four-year periods were 1961 through 1964, and 1965 through 1968. The cash prices used for choice live steers represented the average weekly prices paid for choice steers at Chicago. The pork belly prices, obtained from the National Provisioner, represented 12–14 pound bellies in Chicago. The data were analyzed on the basis of four-year and two-year time periods.

TABLE 1—RANDOM VARIANCES OF CASH PRICES OF LIVE CHOICE GRADE CATTLE AND PORK BELLIES, FOUR-YEAR BASIS[a]

Commodity	Four-Year Period Without Futures Trading	Four-Year Period With Futures Trading	Difference
Pork Bellies	1.040	.62	−.420*
Beef	.086	.041	−.045*

* Difference Significant at 5 percent level.

[a] Original data in cents/pound.

The analysis of the data on the four-year basis indicates that the variance of the random element in live beef prices was reduced from .086 or about 30 cents per cwt. during the four years without futures trading to .041 or about 20 cents per cwt. during the four years with futures trading.[3] The difference between these variances was significant (5%). Similarly, during the four years without futures trading in pork bellies, the variance in the random element was 1.040 or a little more than $1.00 per cwt., while in the four-year period with futures trading the random variance declined to.62 or about 78 cents per cwt. These differences in variance were also significant at the 5 percent level. See Table 1.

The results of the two-year analyses parallel those for the four-year analysis as seen in

[1] See any basic statistics text for the standard-error-difference formula for testing the difference between two variances. An F-test at the 5 percent level was used to determine the significance of the difference.

[2] Four-year periods were considered in order to hold to a minimum the changes that occurred in the structure of the markets concerned.

[3] The variance is measured in (cents/cwt).[2] Thus a variance of .041 [cents/cwt][2] is equivalent to [.20 cents/cwt.][2].

TABLE 2—RANDOM VARIANCES OF CASH PRICES OF LIVE CHOICE GRADE CATTLE AND PORK BELLIES, TWO-YEAR BASIS[a]

Period and Commodity	Period Without Futures Trading	Period With Futures Trading	Difference
Beef			
1	.101	.045	−.560*
2	.073	.041	−.032*
Pork Bellies			
1	.62	.38	−.240*
2	1.29	.85	−.440*

* Difference Significant at 5 percent level.
[a] Original data in cents/pound.

Table 2. In each of the two-year periods considered in live beef, the random fluctuations were significantly lower than in each of the two-year periods without futures trading. Likewise for pork bellies the analysis indicates that for similar years in the price cycle the random fluctuations were significantly lower when there was futures trading than when there was not. On this basis the variances in the random element in pork bellies were .38 and .85 with futures trading compared to .62 and 1.29 in the corresponding periods without futures trading.

It should be noted that in the case of the beef data, the prices represent different years in essentially the same price cycle, while the pork belly data represent years in different cycles. The most logical comparison would be on the basis of corresponding years in successive price cycles. This was not possible in beef. Nevertheless, it was assumed that any changes in random fluctuations that might be related to the upswings or downswings in a cycle would not be major. In any case it would be unrealistic to attribute the nearly 50 percent reduction in random beef price fluctuation to changes in the price cycle.

In summary the evidence suggests that during the time periods considered for pork bellies and live beef the variance of the random element in cash prices for these commodities was significantly lower when futures trading occurred then when it did not.

The question remains "Can this reduction in the variance of the random element be attributed to futures trading?" If so, what logic suggests it? Answers to these questions re not easily found. Part of the difficulty arises from the usual problems associated with the *ceteris paribus* assumption.

The most logical explanation for random variations in prices rests on the degree to which market participants are informed of fundamental supply and demand conditions. The more informed are market participants, the greater the likelihood that the prices arrived at in the market will represent true supply and demand; the less informed they are, the greater the likelihood that prices will deviate from the true equilibrium price, or, the larger the random element will be. On this basis one would expect the fluctuations in price that cannot be explained by fundamental supply and demand conditions to be reduced when market participants are more informed than when they are less informed.

III. *The Information Role of the Futures Market*

Information is the key to competition. Market information has a particularly important place among the factors that determine what is offered for sale and what is demanded, and hence among the factors that determine prices. As markets become more decentralized, information concerning current and future demand and supply conditions must be carefully collected and interpreted.

Commodity future exchanges have been termed clearing centers for information. Information relative to supplies, movements, withdrawals from storage, purchases, current production, general supply and demand conditions, cash and futures prices, and volume of futures trading, is collected, collated, and distributed by the exchange, its members and the institutions such as brokerage houses, which serve the exchange. This information is used not only by current and potential traders in futures, but it is also carefully evaluated by cash market operators.

The existence of futures trading in a commodity should increase the speed with which information is desseminated, the area over which it is disseminated, and the degree of saturation within the area. It should tend to equalize the flows of information to current and potential futures and cash market partic-

ipants.[4] The result should be more informed decision making and prices that are more closely representative of basic supply and demand conditions; prices whose random element is less than it would be without futures trading; price messages that are more sharply defined and less distorted by noise or the random element.

During the time periods considered in this study the only major changes in information flows for these commodities were those resulting from futures trading. Perhaps then, the answer to the first question posed above is, yes, part of the reduction in the variance in the random element can be attributed to the inception of futures trading in these commodities. In answer to the second question, the relationship between the reductions in random price fluctuations and futures trading is explained in part by the improvements in the information flows fostered by futures trading.

This conclusion is necessarily quite tentative. The results of the data analysis are quite significant although they represent a rather small sample of commodities. The impact of the futures market on the market information system and on decisions made by potential and current market participants needs empirical verification.

This conclusion with respect to $V(E_t)$ in the above analysis implies nothing about the effect futures has on $V(S_t)$. Further work needs to be undertaken to determine the effect futures trading has on $V(S_t)$. It may be that $V(S_t)$ may be greater, less than, or the same when futures trading is in effect than when it is not. Further, it may be that futures trading has a different effect on $V(S_t)$ when the commodity considered is storable and seasonally produced than when the commodity is non- or semi-storable and continuously produced.

To further test the hypothesis developed in this paper and other hypotheses on the effect of futures trading on price fluctuations more studies need to be conducted on a wider range of commodities. The recent genesis of futures trading in a number of other commodities should offer ample opportunity for such research in the years immediately ahead.

[4] In the long run both the quality and quantity of information may also be increased.

REFERENCES

P. H. Cootner, ed., *The Random Character of Stock Market Prices*, Cambridge, Mass. 1964.

R. Gray, "Onions Revisited," *J. Farm Econ.*, May 1963, *45*, 273–76.

H. S. Houthakker, "Systematic and Random Elements in Short-Term Price Movements," *Amer. Econ. Rev.*, May 1961, *51*, 164–72.

A. Larson, "Measurement of a Random-Process in Futures Prices," *Food Res. Inst. Stud.*, Nov. 1960, *1*, 313–24.

G. Tintner, *The Variate Difference Method*, Bloomington 1940.

H. Working, "A Random-Difference Series for Use in the Analysis of Time-Series," *J. Amer. Statist. Ass.*, Mar. 1934, *29*, 11–24.

———, "A Theory of Anticipatory Prices," *Amer. Econ. Rev. Proc.*, May 1958, *48*, 188–99.

———, "Price Effects of Futures Trading," *Food Res. Inst. Stud.*, Feb. 1960, *1*, 3–31.

———, "New Concepts Concerning Futures Markets and Prices," *Amer. Econ. Rev.*, May 1961, *51*, 160–63.

———, "Futures Markets Under Reserved Attack," *Food Res. Inst. Stud.*, Feb. 1963, *4*, 13–24.

U.S. Congress. House Subcommittee on Agriculture. "Prohibit Trading in Irish Potato Futures on Commodity Exchange," *Hearings on H.R. 904*, 88th Cong., April 8–10, 1963.

———. Senate. *Hearings Before a Subcommittee of the Committee on Agriculture and Forestry*, 85th Cong., 1st sess., Aug. 12, 1957.

———. Senate. *Hearings Before the Committee on Agriculture and Forestry*, 85th Congress, 2d sess., Mar. 22–26, 1958.

International Differences in Capital-Output Ratios

By Nicholas V. Gianaris*

The allocation of investment to particular industries is a major problem in economic projection and planning. The purpose of this paper is to examine the stability of the sectoral incremental capital-output ratios over time and their differences among countries. It is clear that the usefulness of these ratios in projecting capital requirements will depend on how stable they are, at least over the periods used in development programs.

Section I examines the problem of measuring the incremental capital-output ratio and lists the factors affecting it. Section II offers the empirical results of the behavior (trend and variability) of the ratio in a number of sectors of several countries. Section III deals with the possibility of misallocation of investment even if the ratio is stable, and offers an interpretation of the findings.

I

The incremental capital-output ratio, $ICOR = \Delta K/\Delta Q = I_t/(Q_t - Q_{t-1})$, denotes the number of additional units of capital ΔK, used to produce an additional unit of output ΔQ. In this article, the gross (undepreciated) incremental capital-output ratio (gross domestic fixed capital formation to incremental gross domestic product, $GDFCF/\Delta GDP$, at constant prices) is used because of the difficulty of measuring assets in net terms and because gross investment is itself a major vehicle of technological progress.

The measurement of investment and output for the same period of time seems to be practical, but in some sectors like transportation, electricity, and other public works, the fruits of investment need long periods to be realized and make for changes in the "nominal" *ICOR*s. In such cases, investment requirements may be determined not by *ICOR*s, but other criteria; for example, through the use of engineering studies, evaluation by experts using the experience of other countries, or through extrapolation of previous investment trends. When the fruits of investment have different time paths among sectors, they must be discounted to be comparable. In the present study, no time lag between investment and output is allowed, because no data for any acceptable estimate of the lag are available. Moreover, the exact lag makes less difference when longer than one-year periods are considered.

The factors affecting the sectoral *ICOR*s are so many and so varying that the *ICOR*s can hardly be assumed to be stable in the short run on a priori grounds. Change in factors such as labor, capital accumulation, technological progress, organization, and management do not necessarily offset one another. Changes in policy can have a strong influence on the magnitude and the movement of the sectoral *ICOR*s, especially when they involve shifts toward industries with heavy capital requirements as housing, electricity, and other public works.

Furthermore, the substitution of labor for capital when the price of labor falls in relation to the price of capital, tends to decrease the capital-output ratio.

From the viewpoint of sectoral *ICOR*s, monopolistic practices may have different effects, depending on the particular industries and the stage of development of the country. It would seem that concentration taking place in developed countries leads to technical progress, specialization and economies of scale, better planning and organization and, therefore, to lower *ICOR*s. On the contrary, similar monopolistic firms operating in developing countries are, to a large extent, family-owned and enjoy protection through high tariffs. They may earn com-

* Assistant professor of economics at Fordham University. While the author alone takes the responsibility for his paper, he acknowledges debt to Professors Solomon Fabricant, Hourmouzis Georgiades of New York University and Louis Spadaro and Joseph Cammarosano of Fordham University for their valuable comments.

fortable rates of profit and neglect any search for innovations, better organization, and other competitive practices. Such sleeping monopolies use more capital than competitive firms to produce the same amount of output and make for higher *ICORs*.

Business fluctuations affect the amount of capital used and the amount of output produced. During the early period of an economic upswing the stock of capital is rising but at a lower rate than output, and the *ICOR* is declining. When business is generally contracting, the ratio is increasing as idle capacity is created. Sectors like agriculture and mining which are subject to weather conditions and other external factors are expected to fluctuate more than other sectors such as manufacturing, trade, or banking services.

It has been said by Vinayak Bhatt that if a country is in an early stage of development, a good deal of social overhead capital is needed in transport, power, waterworks, and education. In this stage of economic growth, the *ICOR* in these sectors is high. But after a period of time, when investment in public utilities and unexploited techniques and natural resources take effect, output can be increased with a smaller increase in capital than previously.

It is difficult to say whether the sectoral *ICORs* in developing countries will tend to increase or to decrease over time. Highways, electric power plants, sewer systems, gas pipes, and other public projects, which are necessary for the "infrastructure" of these countries, involve high *ICORs*. But as soon as the required railways, power, or pipelines are established, additional demand can be satisfied with small amounts of additional capital. Indivisibilities in these industries will bring about increasing returns in the later stages of development. Moreover, increase of productive efficiency, improved marketing, and business organization can substantially reduce this ratio. In the long run, the ratio will probably decline since more developed economies generally reveal declining sectoral ratios after the early stages of development, (see R. Goldsmith and C. Saunders and S. Kuznets).

II

This empirical section is directed to the question whether the sectoral *ICORs* are stable enough to justify their use in projecting capital requirements, and whether there are similarities of *ICORs* among different countries.

Sectoral variations of the *ICORs* from year to year can be quite large. As Figures 1 and 2 show, all sectors of the five countries considered had such a wide annual variation that it is impossible for such *ICORs* to be used for further projection. In England, for example, the *ICOR* in manufacturing varied from 3.9 in 1951 to −89.3 in 1956 and 1.4 in 1964; while that of Greece varied from −6.8 in 1952 to 0.5 in 1954 and 2.1 in 1965. Similarly wide annual variations can be observed in all other sectors: transportation, housing, and primarily agriculture, even though a three-year moving average was used in the calculation of *ICORs*. In development planning, however, annual variations do not matter much because longer periods are used.

It is apparent from Figures 1 and 2 that an inverse relationship exists between the growth rates of *GDP* and the *ICORs* for all sectors considered.[1] The main reason for this relationship is that investment is a more stable variable than the rate of growth of *GDP*. This suggests an independent element causing variations in the growth rate of *GDP* which does not simultaneously change the ratio of *GDP* invested. This could be caused by changes in employment or productivity within each country, over time, and across the sample of countries. Capital investment in transportation, electricity, and other durable public works, necessary for the "infrastructure" of a country, takes place more or less in a stable rate, with less regard to the related sectoral short-run rate of growth of output. In many cases, primarily in developing countries, decisions on the sectoral allocation of investment are affected more by political considerations than by the growth rates of *GDP* in the particular sectors.

[1] Statistical data presented by Harvey Leibenstein demonstrate that overall or national *ICORs* are also inversely related to the growth rates of national product.

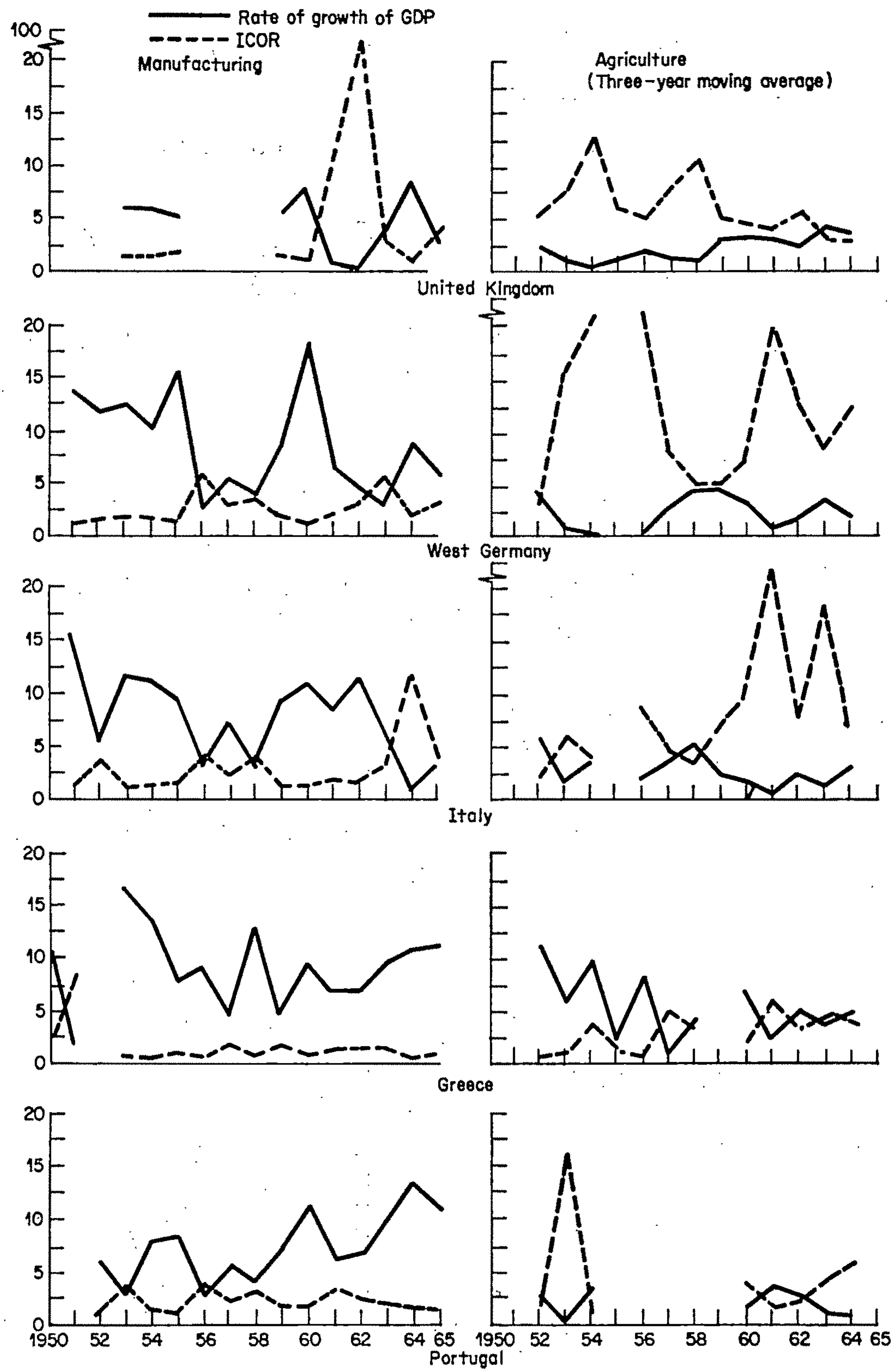


FIGURE 1. ANNUAL RATES OF GROWTH OF GDP AND ICORs IN MANUFACTURING AND AGRICULTURE
SOURCE: CALCULATIONS WERE BASED ON OECD

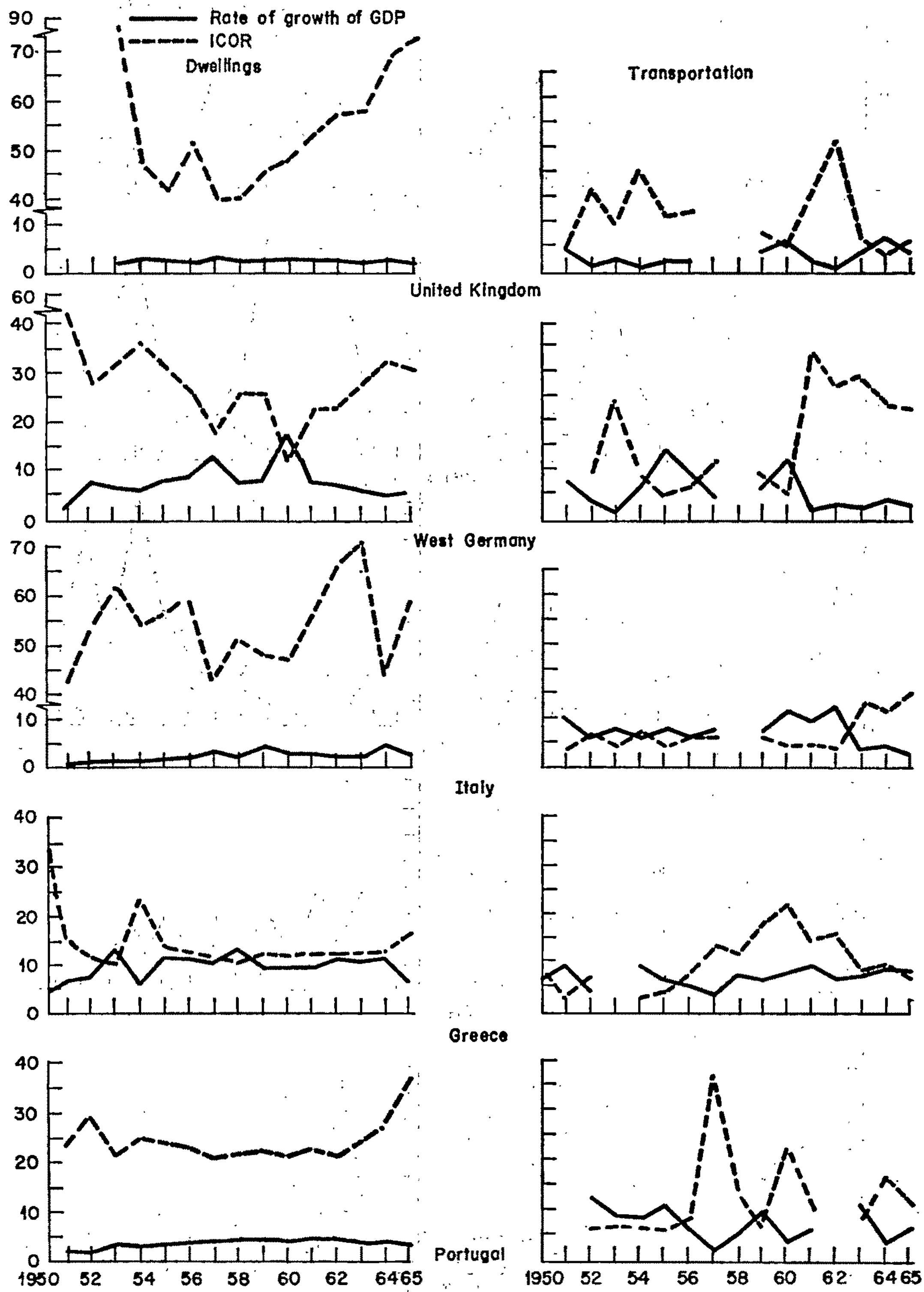


FIGURE 2. ANNUAL RATES OF GROWTH OF GDP AND ICORs IN DWELLINGS AND TRANSPORTATION
SOURCE: THE SAME AS IN FIGURE 1.

The sectoral *ICOR*s are much higher for the years and countries with low growth rates than for those with high growth rates. Thus, Germany and Greece, which had high growth rates in dwellings for 1951–65, had *ICOR*s 25.2 and 13.2, respectively; while England and Italy had low growth rates and *ICOR*s 58.5 and 52.9, respectively.

For purposes of narrowing the extreme variability of year-to-year ratios and because this is the period generally used in projections of economic development, the five-year period is used below for the measurement of the sectoral *ICOR*s. Tables 1 and 2 show the variations of five-year *ICOR*s in eight sectors of the economies of England, Italy, and Germany, representing the developed countries, and Greece and Portugal representing the developing countries, during the three five-year periods of 1951-1965. It is expected, however, that *mutatis mutandis* all other countries would have similar variations in those *ICOR*s.[2]

For the calculation of the five-year *ICOR*s, the total amount of gross investment during the five-year period was divided by the incremental *GDP* during the same period. That is:

$$(1) \qquad ICOR = \frac{\sum_{t=1}^{t=5} I_t}{Q_{t+5} - Q_t}$$

A similar formula was used for the calculation of the 15-year *ICOR*s.

That is:

$$(2) \qquad ICOR = \frac{\sum_{t=1951}^{t=1965} I_t}{Q_{1965} - Q_{1950}}$$

Agriculture and mining presented the largest variations; manufacturing and trade-banking services the lowest; and all other sectors in between. In all three developed countries, the manufacturing five-year *ICOR*s varied between the lowest, 1.1 for Germany, and the highest, 4.0 for England. For Greece and Portugal, the manufacturing *ICOR*s

[2] For additional sectoral and national *ICOR*s, see Appendix, Tables 1 and 2.

TABLE 1—SECTORAL *ICORs* IN FIVE-YEAR TOTALS FOR GERMANY, ITALY, AND THE UNITED KINGDOM, 1951–65.
(*GDP* at factor cost: 1958 prices)

Sectors	Germany				Italy				United Kingdom			
	1951–55	1956–60	1961–65	1951–65	1951–55	1956–60	1961–65	1951–65	1951–55	1956–60	1961–65	1951–65
Agriculture	5.0	6.6	28.6	9.4	3.7	13.4	5.7	6.0	8.6	4.7	5.0	5.6
Mining and quarrying	4.7	7.0	1.0	2.1	6.7	4.6	133.2	8.6	12.9	−7.0	−33.1	−20.3
Manufacturing and construction	1.1	1.8	2.7	1.8	1.6	1.8	2.9	2.2	2.9	4.0	3.6	3.1
Electricity, gas and water	8.0	7.4	9.8	7.9	8.3	7.0	6.3	6.4	14.8	17.1	16.2	16.2
Transportation and communication	6.5	8.6	25.5	11.7	6.2	6.4	6.7	6.6	9.6	11.7	7.5	9.7
Dwellings	33.9	19.7	27.1	25.2	49.3	49.7	57.4	52.9	81.1	47.7	62.1	58.5
Public administration	4.0	4.0	6.2	4.9	8.4	5.4	6.4	6.4	8.3	−6.0	30.0	−168.2
Other services (trade, banking, etc.)	1.1	1.5	3.4	2.1	1.5	2.2	2.8	2.2	3.5	3.3	4.9	3.9
Entire economy	2.4	3.1	5.4	3.7	3.6	4.8	5.8	4.0	5.3	6.8	6.9	6.7

Note: For Germany *GDP* at market prices. For Italy, *GDFCF* of industry for 1963–65 was split in mining (8%) and electricity (17%) according to the average participation of the three last years.
Source: Organization of Economic Cooperation and Development (*OECD*).

TABLE 2—*ICORS* IN FIVE-YEAR SUMS FOR GREECE AND PORTUGAL 1951–1965
(*GDFCF*+Δ*GDP* AT FACTOR COST: 1958 PRICES)

Sectors	Greece 1950–54	Greece 1955–59	Greece 1960–64	Greece 1950–64	Portugal 1951–55	Portugal 1956–60	Portugal 1961–65	Portugal 1951–65
Agriculture	1.3	1.8	3.1	2.1	4.0	4.7	3.5	4.4
Mining and quarrying	2.8	2.3	2.0	2.5	8.8	−7.1	−132.0	28.7
Manufacturing and construction	1.7	1.7	1.5	1.6	1.7	2.2	1.7	1.8
Electricity, gas and water	18.5	10.0	8.9	10.4	7.2	7.9	7.4	7.5
Transportation and communication	5.8	7.1	7.3	7.0	5.8	10.4	13.4	10.3
Dwellings	16.3	14.0	12.2	13.2	25.6	13.3	25.5	23.4
Public administration	5.9	2.2	1.8	2.6	2.7	3.7	1.9	2.5
Other services (trade, banking, etc.)	1.9	1.7	1.8	1.8	1.2	0.9	2.3	1.8
Entire economy	3.3	3.2	3.6	3.4	3.6	3.3	3.2	3.4

Note: For Greece, Public Administration, the period 1951–55 was taken instead of 1950–54. For Portugal, Manufacturing, Transportation, and Entire Economy, the period 1952–5 was taken instead of 1951–5. Prices of 1954 were inflated to 1958.
Source: Same as Table 1.

were more stable, varying between 2.2 to 1.5. In all five countries but Greece, electricity *ICORs* were more stable than transportation and housing. All countries had high *ICORs* in housing, electricity, and transportation. They are capital-intensive sectors, while manufacturing, trade, services and the like use proportionally larger amounts of labor than capital. But this does not mean that it is better to invest more in manufacturing and banking than in trains, buses, or power plants. As will be shown later, in addition to the *ICORs*, it is necessary to know the incremental property share in order to determine whether investment is more productive in manufacturing industry than in housing, transport, and electricity.

The agricultural *ICORs* of the developing countries were lower than in the developed countries, mainly because of the labor-intensive methods of production used in the developing countries. In addition, because of the existing surplus labor in the agricultural sector of the developing countries, a great number of labor hours are devoted to the improvement of land. Making fences, digging ditches, planting new trees, and a host of other farming works are the main operations which the average peasant performs throughout the year, which escape measurement. Capital formation is therefore underestimated, and the ratio is lower than it ought to be.

There is not a great difference in the industrial *ICORs* between developed and developing countries. This can be attributed, probably, to the fact that labor-intensive methods of production in cottage industries in developing countries, which push the *ICORs* downward, are offset by capital-intensive methods of production in modern heavy industries introduced by these countries, which push the *ICORs* upward. On the other hand, advanced technology and better organization used by developed countries bring about economies in scale and make for lower industrial *ICORs*, even though capital intensive methods are used.

Housing or dwellings had very high *ICORs* in all countries, and affected the *ICORs* of the overall economies to a large extent. If this sector were excluded from the entire economy of Greece for example, the overall *ICOR* would have been reduced from 3.4 to 2.5 in 1950–64. The share of housing to total national investment in 1955–64 was 28.6 percent in the United States; 20.5 in England; 26 in France and Italy; and as high as 33.5 for Israel and 37 for Greece. By contrast, developing countries, especially Greece and Israel, had low investment proportions in manufacturing as compared to developed countries, see the United Nations *Yearbook of National Accounts Statistics.*

In addition to the above estimates, a regression analysis of *GDP* on cumulative in-

vestment for the period 1950–65, is employed below to determine the sectoral *ICOR*-slopes and their reliability. Important to this paper are the *ICOR*-slopes, $\Delta K/\Delta Q$, and the variations from the trend, not the absolute figures of capital and output. The regression technique was used for a more reliable measurement of *ICOR*s because the *ICOR* computed from five-year intervals relied too heavily on the terminal year observations.[3] Table 3 shows that in all sectors but agriculture, mining, and public administration, the fit for the regression line was very good. For manufacturing, transportation, dwellings, and trade-banking services, the corrected coefficient of determination, $\overline{R}^2$, was more than 0.935 in all countries. Agriculture, mining, and public administration, however, had such large variations that they hardly can give reliable projections.

The Durbin-Watson statistic is higher than the critical value (1.10 at 5 percent level of significance) in the majority of the countries in agriculture, mining, and manufacturing. In all other cases, however, it falls in the inconclusive range or it is low enough to signify the existence of serial correlation and make the regressions doubtful.

A comparison of the *ICOR*s of the entire period 1951–65 and the *ICOR*s obtained through the regression of *GDP* on cumulative investment (Tables 1–3) shows that the *ICOR*s obtained by these two methods are not far apart from each other. In the first case (the entire period), the *ICOR*s were affected by the output of the first (1950) and the last year (1965), but not by the outputs of the years in between, and, therefore, they were not as representative as in the second case (*ICOR*s obtained through regression of GDP on cumulative investment). Furthermore, this last statistical device, used for long periods, gives values close to the values of the average capital-output ratios.

In the majority of sectors, developing countries had lower *ICOR*s than developed countries. This difference is obvious in agriculture and public administration, primarily, because of the mechanized agricultural process of production and high defense investment in developed countries. Thus, the agricultural *ICOR* in the United States for the period 1950–65 was 5.25 as compared to 9.8 for Germany, 9.6 for Italy, and only 4.7 for Portugal and 2.4 for Greece (Table 3). In electricity and transportation, however, the *ICOR*s of Greece and Portugal were higher than in Germany and Italy. Comparatively speaking, the United States had the lowest *ICOR* in trade-banking and other similar services (1.04) and a low *ICOR* (6.5) in transportation (Table 1 of the Appendix). A more efficient organization and coordination in trade-banking services and the mass production techniques in transport used by the United States might be responsible for the low *ICOR*s.

III

The manufacturing *ICOR* for all countries considered in the empirical part of this study was approximately 2 and that of transportation, 10. This means that for one additional unit of output, 10 units of capital were used in transportation and only 2 units in manufacturing. From the viewpoint of the marginal rate of return of capital, this comparison could well be misleading because different kinds of output in each country or in each sector require different quantities of the other factors of production.[4] Also, investment in one sector can permit output increases in other sectors.

The sectoral differences in the *ICOR*s can be due primarily to the difference in the con-

[3] Hollis Chenery and Alan Strout used national *ICOR* trends of six observations (1957–62) for projections of 1975. In this paper the emphasis is placed on sectoral *ICOR*s and the *ICOR*-slopes are based on sixteen observations (1950–65).

[4] John Adler, for example, concluded that "... where income and rates of investment are low, the yields of investment in the form of additional output is proportionately greater than in countries with higher incomes and higher rates of investment." To support his conclusion, he compared the low net *ICOR* of India, 2.1 for 1948-49/1953-54, with that of Canada, the United States, and Western Europe which was approximately 4.1 for the same period. Also Jeffrey Nugent, discussing Andreas Papandreou's conclusions, seems to imply that more investment must be channeled to manufacturing because the capital-output ratio in this sector is low.

TABLE 3—SECTORAL REGRESSION ANALYSIS BY COUNTRY, 1950–65 (1958 PRICES)

Countries	Intercept[a] a	Regression Coefficient b	T Value	*ICOR*-Slope	$\bar{R}^2$	Durbin-Watson Statistic
Agriculture						
United States	19.81	0.0816 (0.0143)	5.70	12.50	0.677	2.041
United Kingdom	0.73	0.1697 (0.008)	22.16	5.88	0.970	1.063
Italy	2.51	0.1042 (0.011)	9.38	9.62	0.853	1.957
Germany	14.24	0.1017 (0.009)	10.73	9.83	0.884	1.760
Greece	18.24	0.4107 (0.057)	7.18	2.44	0.771	1.238
Portugal[b]	14.36	0.2135 (0.031)	6.82	4.68	0.778	1.972
Mining and quarrying						
United Kingdom	0.76	−0.0699 (0.011)	−6.35	−14.29	0.735	0.543
Italy	0.07	0.1326 (0.019)	7.00	7.52	0.762	0.344
Germany	6.21	0.2331 (0.060)	3.91	4.29	0.487	1.273
Greece	0.17	0.4474 (0.0144)	31.16	2.23	0.985	1.199
Portugal[b]	0.41	−0.0373 (0.007)	−5.33	−27.03	0.667	1.574
Manufacturing and construction						
United Kingdom	6.50	0.2745 (0.013)	20.88	3.64	0.967	1.417
Italy	2.87	0.4719 (0.0159)	29.63	2.12	0.983	0.579
Germany	57.23	0.5484 (0.024)	22.87	1.83	0.972	0.488
Greece	4.64	0.8791 (0.021)	42.27	1.14	0.992	1.177
Portugal[b]	12.97	0.5137 (0.010)	51.62	1.95	0.995	1.261
Electricity, gas and water						
United Kingdom	0.34	0.0630 (0.001)	58.00	15.87	0.996	1.322
Italy	0.17	0.1550 (0.005)	31.48	6.45	0.985	0.668
Germany[b]	2.35	0.1285 (0.003)	41.45	7.78	0.992	1.505
Greece	0.30	0.0918 (0.003)	33.62	10.89	0.987	0.611
Portugal[b]	0.72	0.0554 (0.008)	7.29	18.05	0.801	1.292
Transportation and communication						
United Kingdom	1.41	0.0875 (0.006)	14.88	11.43	0.936	0.583

TABLE 3—SECTORAL REGRESSION ANALYSIS BY COUNTRY, 1950–65 (1958 PRICES)

Countries	Intercept[a] a	Regression Coefficient b	T Value	*ICOR*-Slope	$\bar{R}^2$	Durbin-Watson Statistic
Transportation and communication						
Italy	0.70	0.1732 (0.005)	33.47	5.78	0.987	1.027
Germany	10.47	0.0885 (0.005)	14.71	11.30	0.935	0.476
Greece	4.16	0.0640 (0.003)	18.65	15.63	0.959	0.624
Portugal[b]	2.21	0.0868 (0.004)	20.64	11.52	0.970	0.885
Dwellings						
United Kingdom	0.66	0.0192 (0.001)	34.62	52.08	0.988	0.476
Italy	0.69	0.0189 (0.0002)	92.04	52.91	0.998	0.894
Germany	2.49	0.0427 (0.001)	41.61	23.26	0.991	0.612
Greece	1.51	0.0962 (0.006)	14.71	10.42	0.935	1.288
Portugal[b]	1.66	0.0438 (0.001)	67.45	22.83	0.997	0.786
Public administration						
United Kingdom	1.31	−0.0522 (0.015)	−3.43	−19.23	0.418	0.439
Italy	1.36	0.1427 (0.005)	29.75	6.99	0.983	0.473
Germany	12.46	0.2128 (0.006)	32.84	4.79	0.986	1.022
Greece	4.66	0.6737 (0.137)	4.93	1.48	0.608	0.447
Portugal[b]	2.13	0.4453 (0.029)	15.52	2.25	0.949	0.746
Other services (Trade-banking, etc.)						
United Kingdom	5.05	0.2744 (0.008)	32.25	2.64	0.986	0.591
Italy	2.41	0.4016 (0.006)	63.20	2.49	0.996	0.925
Germany	37.19	0.4869 (0.026)	18.43	2.05	0.958	0.341
Greece	10.15	0.6254 (0.034)	18.61	1.60	0.958	1.250
Portugal[b]	9.97	0.5467 (0.019)	29.38	1.83	0.985	1.084
Entire economies						
United States	345.70	0.19490 (0.0082)	23.82	5.13	0.974	0.971
United Kingdom	18.55	0.18233 (0.0044)	41.42	5.48	0.991	1.689
Italy	11.90	0.24712 (0.0053)	46.92	4.05	0.993	0.629

TABLE 3—SECTORAL REGRESSION ANALYSIS BY COUNTRY, 1950–65 (1958 PRICES)

Countries	Intercept[a] a	Regression Coefficient b	T Value	*ICOR*-Slope	$\bar{R}^2$	Durbin-Watson Statistic
Entire economies						
Germany	137.57	0.27952 (0.0105)	26.59	3.57	0.979	0.566
Greece	51.76	0.32572 (0.0117)	27.94	3.07	0.981	0.886
Portugal[b]	45.50	0.33079 (0.0056)	59.04	3.03	0.996	1.249

[a] The constant term a (the *GDP*-intercept) is expressed in billions of national monetary units. $ICOR=1/b=1/IOCR$. Regression equation: $Y=a+bX$; where Y=Gross Domestic Product and X=Cumulative Investment.

[b] For all sectors in Portugal and for electricity in Germany, the data of 1950–64 were used.

Source: The same as Table 1.

tributions of the services of several factors of production to output. Thus, the *ICORs* in housing, transportation, and electricity are higher than those in agriculture, mining, manufacturing, and trade services, not because capital is less productive in the sectors of the first category, but because the relative contribution of the services of capital is higher. In other words, labor participates more in agriculture, mining, manufacturing, trade, and banking than in housing, transportation, and electricity. To increase output by a given amount would require less capital in the former industries than in the latter, but it would require more labor.

The *ICOR* is not a guide to the profitability or desirability of investment projects because it does not measure the marginal rate of return on capital. In order to determine in which country, sector, or particular industry, capital is more productive, the property share must be considered along with the capital-output ratio. Using the well-known tautological relationship between the property share and the rate of return of capital and the capital-output ratio,[5] all in an incremental sense,

$$\frac{\Delta R}{\Delta Q}=\frac{\Delta R}{\Delta K}\cdot\frac{\Delta K}{\Delta Q}, \tag{3}$$

we have:

$$r=\frac{\Delta R/\Delta Q}{ICOR}=\frac{\Delta R/\Delta Q}{\Delta K/\Delta Q}=\frac{\Delta R}{\Delta K} \tag{4}$$
$$=\text{Marginal Rate of Return}$$

For example, assume $ICOR=4$, and $\Delta R/\Delta Q=20\%$. We have $r=20\%/4=5\%$ = Marginal Rate of Return, where R is property income, Q output or income, and K capital.

If the incremental property share, $\Delta R/\Delta Q$, for a sector such as manufacturing is the same in two countries, then (*ceteris paribus*) the marginal rate of return is higher in the country with the lower *ICOR* and lower in the country with the higher *ICOR*. In this case, the *ICOR* and the marginal rate of return are inversely related. The same formula can be applied not only to the same sector of different countries, but also to different sectors of the same country. Thus, if the manufacturing *ICOR* were stable and equal to 2 and the incremental property share in manufacturing $\Delta R/\Delta Q$ were 10 percent, the marginal rate of return in this sector would be 5 percent ($10\%/2=5\%$). If the *ICOR* in transportation were also stable and equal to

[5] For a stimulating discussion between Tibor Scitovsky, Edward Denison, and Franco Modigliani, on the constancy of the three ratios, property income share (R/Q), the rate of return of capital (R/K), and the capital-output ratio (K/Q), see Scitovsky; see also the related comment by Kravis in his review (p. 189).

10 and the incremental property share in this sector were 60 percent, the marginal rate of return in transportation would be 6 percent (60%/10=6%), which is higher than that of manufacturing, even though the *ICOR*s were stable in both sectors. Therefore, the Medusa-like fascination of the constancy of the capital-output ratio might mislead policy makers to project investment uneconomically, from the viewpoint of marginal rate of return.

The practical policy-makers, however, would be interested more in social or economy-wide productivity of capital rather than pushing development in one particular sector merely because the marginal rate of return of capital for that sector happened to be high. The consideration of the incremental property share along with the *ICOR* is not necessarily a serious restriction on the usefulness of the *ICOR* when used simply to determine capital requirements. But it is a severe qualification if capital "requirements" so determined are taken to suggest investment that may be expected to be "profitable."

From the previous analysis it is obvious that the *ICOR*s in sectors like agriculture, mining, and public administration are not stable enough for reliable economic projections. In manufacturing, electricity, and trade-banking services where the variation was lower, however, extrapolated trends could be used for projection with some degree of caution. Countries with moderately good historical data and reasonably good statistical fit can use sectoral *ICOR*-trends for projecting capital requirements. ***Mutatis mutandis***, these *ICOR*s can also be used by extrapolation to countries without historical data. Some flexibility and some alternative sets of estimates with different possible targets might be helpful in economic projections. Deviations are always possible and projected *ICOR*s must be subject to modifications during the implementation of the development plans. Much more needs to be done to prepare the way for better estimations. More research and more reliable data, covering not only the capital-output ratios but also the property shares are needed.

Appendix

Table 1—Incremental Capital-Output Ratios in Several Countries by Sector, 1955–1964[c]

Sectors	Belgium	France	Netherlands	United States	USSR[a]	Israel[b]	Korea[b]
Entire economy	5.23	4.66	5.65	5.31	3.87	3.39	2.39
L. Agriculture	3.56	9.05	13.35	−13.69	10.13	6.16	0.72
2. Mining and Quarrying	−1.24	7.93		20.80			0.68
3. Manufacturing	3.46			2.89			3.05
4. Construction	2.37	2.54	4.28		1.96	2.37	0.65
5. Electricity and water	7.09	13.94		14.13			11.71
6. Communication and Transportation	9.50	11.11	12.47	6.53	5.09	6.55	5.99
7. Dwellings	85.00	51.43	25.97	58.59	11.65	15.94	11.54
8. Public administration	6.61	4.42	8.39	4.68			−2.50
9. Other services (Trade-banking, etc.)	3.16	1.81	2.61	1.04	6.98	1.46	1.73

[a] 1959–64 *CDFCF* at 1955 prices; Net Material Product (*NMP*) instead of *GDP*.
[b] 1956–64.
[c] Gross Domestic Product at 1958 Factor Cost; Gross Domestic Fixed Capital Foundation at 1958 Market Prices.
Source: See Table 1 in text.

TABLE 2—NATIONAL *ICORs* (GROSS DOMESTIC FIXED CAPITAL FORMATION ÷ INCREMENTAL GROSS DOMESTIC PRODUCT) IN FIVE-YEAR SUMS, CONSTANT 1958 MARKET PRICES

Countries	1951–55	1956–60	1961–65	1951–65
I. Developed Countries				
1. Austria	3.4	3.5	6.2	4.7
2. Belgium	5.3	5.6	4.2	5.2
3. Canada	5.2	7.6	4.0	5.3
4. Denmark	4.0	4.6	4.6	4.3
5. France	3.9	3.8	4.4	3.9
6. Germany, W.	2.4	3.0	5.4	3.7
7. Italy	3.3	3.3	4.8	4.1
8. Netherlands	4.1	5.9	5.4	5.1
9. Norway	4.8	8.2	5.9	7.6
10. Sweden	5.9	6.1	4.8	5.6
11. United Kingdom	4.5	5.9	5.7	5.3
12. United States	4.3	6.9	3.7	4.8
II. Developing Countries				
1. Ceylon	2.4	3.5	4.2	3.4
2. Chile	5.4	3.5	4.2	3.4
3. Colombia	4.6	4.7	4.0	4.4
4. Greece	2.1	4.0	2.4	3.1
5. Guatemala	3.8	2.4	1.8	2.3
6. Honduras	7.2	3.0	3.1	3.7
7. Israel	2.3	3.1	2.8	2.8
8. Nigeria	2.0	6.8	3.3	3.3
9. Philippines	1.0	2.0	2.4	1.8
10. Portugal	3.3	3.0	3.0	3.0
11. Thailand	1.9	2.6	2.9	2.6
12. Venezuela	3.0	3.9	3.2	3.5

Note: For Portugal and Thailand, 1952 was the first year instead of 1951. For Nigeria, 1963 was the last year instead of 1965; *GDP* in factor cost.

Source: Calculations were based on:

i) Organization of Economic Cooperation and Development (*OECD*), *Statistics of National Accounts*, 1950–61, 1955–62, 1956–63

ii) United Nations, *Yearbook of National Accounts Statistics*, various issues.

REFERENCES

J. H. Adler, "World Economic Growth: Retrospect and Prospects," *Rev. Econ. Statist.*, Aug. 1956, *38*, 273–85.

W. Beckerman, et al. *The British Economy in 1975*, Nat. Ins. Econ. Soc. Research, Cambridge 1965, pp. 253–63.

V. Bhatt, "Some Further Notes on Aggregate Capital-Output Ratio," *Indian Econ. J.*, April-June, 1964, *11*, 383–99.

G. Borts and J. Stein, *Economic Growth in a Free Market*, New York 1964.

H. Chenery and A. Strout, "Foreign Assistance and Economic Development," *Amer. Econ. Rev.*, Sept. 1966, *56*, 679–733.

E. Domar, "The Capital-Output Ratio in the United States: Its Variation and Stability," in F. Lutz, ed., *The Theory of Capital*, New York 1961, 95–119.

R. Goldsmith and C. Saunders, ed. *The Measurement of National Wealth*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Ser. VIII, Chicago 1959, pp. 116, 224, 306.

I. B. Kravis, Book Review of, *Amer. Econ. Rev.*, Mar. 1965, *55*, 187–93.

S. Kuznets, *Capital in the American Economy*, Nat. Bur. Econ. Res. Princeton 1961, pp. 205, 217, 250.

———, "Quantitative Aspects of the Economic Growth of Nations," *Econ. Develop. Cult. Change*, VII, N.3, Part II, July 1960; IX. No. 4, Part II, July 1961.

H. Leibenstein, "Incremental Capital-Output Ratios and Growth Rates in the Short Run," *Rev. Econ. Statist.*, Feb. 1966, *48*, p. 24.

J. Nugent, "Economic Thought Investment Criteria, and Development Strategies in Greece—A Postwar Survey," *Econ. Develop. Cult. Change*, April 1967, *15*, p. 334.

A. Papandreou, *A Strategy for Greek Economic Development*, Athens 1962.

E. Phelps, ed., *The Goal of Economic Growth*. New York 1962.

W. Reddaway, *The Development of the Indian Economy*. London 1962.

T. Scitovsky, "A Survey of Some Theories of Income Distribution," Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Ser. XXVII, Princeton 1964.

P. Sen, "Use of the Capital-Output Ratio in Economic Planning," *Indian Econ. Rev.*, Feb. 1960, *5*, 23–31.

R. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

J. Stein, "Economic Growth in the West," *Economica*, Feb. 1965, *32*, 80–87.

J. Vanek, *Estimating Foreign Resource Needs for Economic Development: Theory, Method and Case Study of Columbia*. New York 1967.

Organization for Economic Cooperation and Development (O.E.C.D.) *Statistics of National Accounts*, 1950–61, 1955–62, 1956–65, Paris.

United Nations, *Programming Techniques for Economic Development*. New York 1960, p. 11.

———, *Yearbook of National Accounts Statistics, various issues.*

A Fisherian Approach to Trade, Capital Movements, and Tariffs

By MICHAEL CONNOLLY AND STEPHEN ROSS*

The purpose of this study is to introduce the element of time into the analysis of optimal taxation of international capital movements. As with Irving Fisher, in his *Theory of Interest*, the "... supply and demand we have to deal with are ... the supply and demand of future income," and we shall interpret a rate of interest as "... that sort of price which links one point of time with another point of time in the markets of the world" (pp. 32–33). The analysis stresses the formal identity between capital theory, involving exchange over time, and trade theory, involving exchange at the same point of time and argues that, as a consequence, the theory of the optimum tax on capital movements may be regarded as a branch of optimal tariff theory.

The correspondence between trade theory and capital theory has appeared in many guises throughout the past century, and was recognized at least as early as 1907 by Fisher. His classic graphical apparatus anticipates much of the later opportunity cost approach to international trade, as well as the more recent work in general equilibrium encompassing both theories.[1] In this latter work, time is broken up into a (finite) number of successive time periods, and the same physical item available at different time periods and locations is treated as a different commodity. Future production and consumption possibilities are assumed to be perfectly known today and economic agents acting as price takers engage in markets in which prices for delivery in future periods are quoted, and, under certainty, will correctly reflect future scarcity values.

As an immediate consequence of this approach, optimum tariff theory is no less applicable when trade over time is included with trade at a given point in time. Prices quoted in the present reflect marginal rates of substitution not only between present commodities, but also between future commodities, and between present and future commodities. A country then stands to gain by taxing exchange over time in the same manner as it may in taxing exchange at the same point in time. Indeed, it has the same purpose—to influence the terms of trade. That the terms of trade over time are called interest rates, or that a lender is treated as an exporter and a borrower as an importer of present income makes little difference. The resulting optimum taxes on trade are simply those taxes which equate the marginal rates of transformation through trade at the same point in time, and over time, with the marginal rates of transformation in domestic production and consumption.

This is a somewhat different approach to the problem of optimal taxation of capital movements than that recently pursued by Murray Kemp and Ronald Jones. The major difference is the explicit role played in our analysis by the element of time—of waiting—which allows for pure borrowing or lending, in addition to the financing of imported capital goods by current exports. In a later section we will relate the two approaches.

I. *Optimal Taxes on Exchange*

Consider a world composed of two countries—the home country and the foreign one—and two goods, bread and wine. The world lasts for only two periods, the present

* The authors, assistant professors of economics at Harvard University and the University of Pennsylvania, respectively, are grateful to Richard Caves, Gottfried Haberler, Thomas Horst, Harry Johnson, Murray Kemp, Robert Mundell, Lawrence Officer, Trevor Underwood, and Thomas Willett for helpful suggestions. They would also like to express their thanks to workshops at the University of Iowa, the University of Chicago, and Harvard University where earlier versions of this paper were presented.

[1] On these points, see Paul Samuelson and Gerard Debreu.

and the future, and there are thus a total of four commodities; present bread and future bread, and present wine and future wine. Perfect futures markets exist and individuals make contracts payable in the present for delivery in both periods. We represent foreign prices in the following table:

	Bread	Wine
Present	P_1^0	P_2^0
Future	P_1^1	P_2^1

where the superscripts, 0 and 1, distinguish presently quoted prices for present and future delivery, and the subscripts, 1 and 2, denote bread and wine respectively. The prices are given in terms of some abstract unit of account, say *bancors*, that will play no role in the subsequent analysis; only relative prices are of importance. The prices of goods for future delivery, P_1^1 and P_2^1, can be thought of as the present values in *bancors* of one unit of the good promised for delivery in the future. Since we also assume that the inhabitants live in a world of certainty and enjoy perfect foresight there can be no divergence between the presently quoted price of bread in terms of wine for future delivery, P_1^1/P_2^1, and the rate of exchange that will prevail in the future period.

Domestic imports of the four commodities can be displayed in a similar table with the same notation:

	Bread	Wine
Present	Z_1^0	Z_2^0
Future	Z_1^1	Z_2^1

The exchange possibilities here are numerous: for example, the home country could import both wine and bread in the present, and export both in the future. This would be a clear case of borrowing today with repayment tomorrow. Alternatively, negative values of Z_1^0 and Z_2^0 indicating exports of both commodities in the present, combined with positive values of Z_1^1 and Z_2^1 indicating future imports, would represent a clear case of lending. What can be said, however, of less obvious cases? For instance, suppose the country imports wine now and exports it in the future, and, at the same time, exports bread now and imports it in the future. Can the country be said to borrow wine and lend bread? Not in a meaningful sense, because borrowing and lending cannot in general be identified with trade in a particular commodity. A country may well be exporting equipment and machinery, and if the value of its exports falls short of the value of its imports, it is incurring debt rather than lending. That is to say, what constitutes borrowing or lending over time depends upon whether or not imports pay for exports in the present, and is, therefore, purely a net concept. In general, $P_1^0Z_1^0+P_2^0Z_2^0$ is the value of a country's debt if it is positive and of its loans if it is negative. If it is zero, the country is neither a debtor nor a creditor.

If a country is borrowing, it is, in effect, selling claims on future goods equal to the value of its debt, and if it is lending, it will be demanding such claims. In other words, while the terms Z_1^1 and Z_2^1 were defined as future imports of each commodity, they may also be thought of as indicating purchases of securities by the home country in the present, each security entitling the holder to a future import of one unit of the commodity in which it is denominated. Clearly, then, $P_1^1Z_1^1+P_2^1Z_2^1$ represents not only the value of commodity imports in the future, but also the value of imports of such claims in the present. It thus follows in direct analogy to the static theory where all markets clear (and as a consequence of Walras' Law) that we must have trade in balance, i.e.:

$$(1)\quad F \equiv P_1^0Z_1^0 + P_2^0Z_2^0 + P_1^1Z_1^1 + P_2^1Z_2^1 = 0$$

Equation (1) simply expresses the relation that the present value of the balance of trade must be zero. Alternatively, if we interpret $P_1^0Z_1^0+P_2^0Z_2^0$ as the balance on current account and $P_1^1Z_1^1+P_2^1Z_2^1$ as the balance on capital account, then it states that the two must cancel.

Following J. de V. Graaff, we will call F the *foreign trade transformation function* facing the home country.[3] On the assumption

[3] We have been careful to adhere closely to Graaff's notation since the argument of the remainder of this

that each of the prices is a differentiable function of all the imports we may differentiate F with respect to each of the import terms to derive the marginal rates of transformation through foreign trade facing the home country:

$$(2)\quad -\frac{dZ_j^\tau}{dZ_i^t} = \frac{\partial F}{\partial Z_i^t} \Big/ \frac{\partial F}{\partial Z_j^\tau} = \frac{P_i^t(1+a_i^t)}{P_j^\tau(1+a_j^\tau)}; \qquad \begin{matrix} t, \tau = 0, 1 \\ i, j = 1, 2 \end{matrix}$$

$$(3)\quad \begin{aligned} a_i^t &\equiv \sum_{(\alpha,k)} \frac{\partial P_k^\alpha}{\partial Z_i^t} \frac{Z_k^\alpha}{P_i^t} \\ &= \sum_{(\alpha,k)} \eta_{it}^{k\alpha} \frac{P_k^\alpha Z_k^\alpha}{P_i^t Z_i^t}, \qquad \begin{matrix} i, k = 1, 2 \\ t, \alpha = 0, 1 \end{matrix} \end{aligned}$$

and $\eta_{it}^{k\alpha}$ is a quantity elasticity, i.e., the elasticity of the price P_k^α with respect to the imports Z^t. From (3) the a^t terms are the sums of the (foreign) quantity elasticities each weighted by the ratio of the total value of trade in that item to the value of trade in Z_i^t and they express the marginal cost to the home country of changes in the terms of trade induced by imports of the i^{th} good in the t^{th} period. Consequently, $P_i^t(1+a_i^t)$ is the total marginal cost now to the home country of increasing its imports of good i during period t or, analogously, the present marginal export revenue if Z_i^t is negative. If the foreign transformation function is infinitely price elastic in all goods and time periods, i.e., if the home country can have no effect upon prices, then all the a_i^t terms are zero and, as in the static case, equation (2) expresses the familiar equality of the rates of transformation and the relative prices.

It is well known that in the absence of externalities, Pareto optimality requires the equality of the domestic marginal rates of substitution in consumption, *DRS*, and the domestic marginal rates of transformation in production, *DRT*, and in an open economy both must be brought into equality with the marginal rates of transformation through trade, *FRT*. These conditions are unaltered when exchange over time is included along with exchange at a given moment of time. On the assumption that the domestic price ratios π^t/π_j^τ reflect the *DRS* and the *DRT*, these efficiency conditions become:

$$(4)\quad \frac{\pi_i^t}{\pi_j^\tau} = \frac{P_i^t(1+a_i^t)}{P_j^\tau(1+a_j^\tau)}; \qquad \begin{matrix} t, \tau = 0, 1 \\ i, j = 1, 2. \end{matrix}$$

The necessary conditions for Pareto optimality may then be satisfied by levying an *ad valorem* tax equal algebraically to a_i^t on each good i traded at time t. For commodities traded in the present, the tax is expressed as a percentage of the foreign commodity price. For trade in contracts to future commodities, the tax is expressed as a percentage of the current foreign price of a claim entitling the holder to future delivery of one unit of the commodity in which it is denominated. Notice also that the taxes are payable in the present, and not necessarily at the time when the good crosses the border. If the taxes were payable upon delivery, they would be costed up at an appropriate interest rate.

In our example this procedure defines a system of four tariffs some of which, due to cross effects, may be subsidies.[3] Barring such cross effects, the a_i^t term for exports is the reciprocal of the foreign elasticity of demand, and is therefore typically negative, and for imports it is the reciprocal of the foreign elasticity of supply and hence positive, im-

section relies upon his derivation of optimum tariffs. The reader may recall the definition he adopts as $\Sigma P_i Z_i + K = 0$, where the summation represents the trade balance, and K foreign lending. He assumes K is constant, given independently of prices, but we join him in saying that "... ; The details, but not the substance of the argument is affected if this assumption is relaxed" (p. 52). Indeed, showing this is part of our exercise. Further, we assume that the foreign country does not retaliate by raising tariff barriers or other restrictions of its own. For perfectly applicable propositions concerning optimum tariffs and retaliation, see H. G. Johnson.

[3] Since only relative prices are of importance it would be possible to satisfy the efficiency conditions with only three taxes. This would amount to choosing one of the goods as a numeraire. We might also mention that convexity of the production and consumption sets is sufficient to guarantee that a global maximum is reached by this system of taxation.

plying a tax in both instances. However, in the general case, the own effects may be outweighed by substitutability and complementarity relationships so that subsidies on trade in some of the goods may be warranted. For instance, it might be worthwhile to subsidize exports of wine in the present if this makes for cheaper wine imports in the future. From equation (3) it can be seen that this situation is most likely to arise if the value of present wine exports is small relative to the value of future wine imports. This example appears to illustrate an obvious case of present sacrifice, a subsidy on foreign lending, compensated for by future reward. It is possible, however, for the total value of present wine exports to fall short of the total value of present bread imports. Since this would entail a deficit on current account, we could now interpret the subsidy as (in part) a tax on additional borrowing whose function is to reduce the current deficit, thereby lowering future repayments. This example serves to illustrate possible pitfalls inherent in identifying capital movements solely with trade in a particular good. The question of the direction and the magnitude of capital movements in this framework revolves mainly around the question of whether the country is a *net* borrower or lender.[4]

II. *Interest Rates*

In principle, the task of finding expressions for optimum taxes on exchange is finished, we need not dwell upon matters such as interset rates: the equivalent information is contained in the terms of trade over time. We will do so, however, to aid in the interpretation of the formulas obtained above. In particular, we should like to examine the relationship that domestic interest rates bear; first, to foreign interest rates, and second, to one another. We have two standards in which to express interest rates—in bread, and in wine. Letting r_1 and r_1^* denote, respectively, the foreign and domestic rates of interest in terms of bread, the precise relationship between the own rates of exchange over time and own interest rates is $P_1^0/P_1^1 = 1+r_1$ and $\pi_1^0/\pi_1^1 = 1+r_1^*$. Thus, from (4) we have that:

$$\text{(5)} \qquad \frac{1+r_i^*}{1+r_i} = \frac{1+a_i^0}{1+a_i^1}\,; \quad i = 1, 2$$

It follows that the domestic rate of interest in good *i* exceeds the foreign one if, and only if, the tax rate on current delivery algebraically exceeds the tax rate on contracts for future delivery.[5]

Letting $\delta = (P_1^1/P_2^1 - P_1^0/P_2^0)/(P_1^0/P_2^0)$, the percentage rate of appreciation of bread in terms of wine, it is easy to verify that:

$$\text{(6)} \qquad r_2 - r_1 = \delta(1 - r_1)$$

which is simply an expression of the Fisherian rule that, ignoring δr_1 as negligible, "... ; Two rates of interest in ... two diverging standards will, in a perfect adjustment, differ from each other by each other by an amount equal to the *rate* of divergence between the two standards" (Fisher, 1930, p. 39).

Clearly, the entire argument of Section I could be formulated in a very straightforward manner in terms of interest rates by the use of the above relationships. We will do so, however, for a very specific example where trade consists solely of the borrowing or lending of a single good. The application of the optimum tax to the two-period Fisherian world provides an illuminating analysis of the general problem of the optimal regulation of international capital movements on an aggregative level. Consider a country which borrows imports, exporting securities or claims on its future income. In pursuing an optimum tax policy, it must bring the marginal cost of borrowing measured in terms of its future exports into equality with the corresponding domestic marginal rate of substitution. From the general optimum tariff formulas, this requires the im-

[4] It should be clear that the formulas and the results of this section generalize simply to the case of an arbitrary number of goods, *n*, and of time periods, T.

[5] This relation holds for the region $a_i^t > -1$. If any of the a_i^t terms fell short of -1, it would pay the home country to *increase* taxes on trade in that good because more imports could be had for fewer exports. This would correspond to the inelastic region of the foreign offer curve in the two-good case, a region in which the home country, as a monopolist, should not operate.

position of an *ad valorem* tax on foreign borrowing of $1/e_{sp}$, where e_{sp} is the elasticity of the foreign supply of loans of present goods with respect to the international terms of trade over time.[6] This is simply the classical two-good trade result where the optimal tax is set equal to the reciprocal of the foreign elasticity of supply with respect to the terms of trade.

Recalling that loans by foreigners to the home country, Z, are equal to the amount of saving done by foreigners, S, minus the amount that they invest at home, I, and that both depend upon the terms of trade over time, we can derive a formula for the optimal tax in terms of the elasticities of savings and investment in the foreign country with respect to the rate of interest. Note first that since $Z=S-I$:

$$\epsilon_{sp} = (S/Z)\epsilon_{sp} + (I/Z)\epsilon_{ip}, \tag{7}$$

that is, the foreign elasticity of the supply of loans equals a weighted sum of the elasticities of saving and investment with respect to the terms of trade over time. Using the identities between price and interest elasticities, $\epsilon_p=((1+r)/r)\epsilon_r$, where r is the international rate of interest, formula (7) can be written in terms of interest elasticities. The optimal tax on borrowing is therefore given by:

$$\tau=\frac{1}{\{(1+r)/r\}[(1+1/\mu)\epsilon_{sr}+(1/\mu)\epsilon_{ir}]} \tag{8}$$

where $\mu=Z/I$, the ratio of the home country's borrowings to total investment by foreigners.

Formula (8) supports a variety of intuitively plausible propositions. All other things held constant:

1. The smaller is μ, the amount foreigners invest in the home country expressed as a percentage of their domestic investment, the smaller will be the optimal tax on borrowing.
2. The more inelastic the foreign schedules of saving and investment with respect to the interest rate, the greater will be the resulting tax.
3. The greater the final equilibrium interest rate, the greater will be the tax.

It must be emphasized that the above are *ceteris paribus* propositions designed solely to isolate the influence of the respective variables as though they could be altered independently.

Purely as an illustrative exercise, we can substitute some hypothetical figures into formula (8). Suppose, for example, that interest rates are 10 percent, and that the *price* elasticities of saving and investment equal unity, so that their interest elasticities are 1/11. Also, let borrowings be 1/6 of total investment made by foreigners. The optimum tax in this instance would be 1/13 or about 8 percent of the loan. If the country cut its dependence on foreign loans in half, the optimum tax rate would fall to 1/25, or 4 percent, reflecting its smaller size in the capital market.

If the home country is lending abroad a similar derivation yields the optimal tax on foreign loans, shown as equation (9) below, which may be analyzed in a manner analogous to that of the tax on domestic borrowing, where μ^* is $-\mu$; i.e., the ratio of the home country's *loans* to total investment by foreigners.

In this form, the optimal tax formulas lend themselves quite easily to interpretation, but clarity has been gained at the expense of the information contained in the general formula (2). In particular, the many possibilities of substitutability and complementarity at the same point in time, and over time in the general case have given way to the simple two good case of substitution over time. Consequently, the case of subsidizing current lending that we examined in Section I cannot arise in this context.

[6] By virtue of the assumption that the foreign country imposes no tariff, its domestic price is identical to the international price.

$$\tau' = \frac{1}{\{(1+r)/r\}[(1/\mu^*)\epsilon_{ir} + (1/\mu^* - 1)\epsilon_{sr}] - 1} \tag{9}$$

In both instances, the problem of optimum taxes on capital movements, including lending and borrowing, has been treated as a branch of the theory of optimum tariffs. Further, as in tariff theory, the tax depends directly upon the relative monopoly power of the country in world capital markets.

III. *The Kemp-Jones Approach*

The purpose of this section is to relate the above approach to the work of previous writers in the field, and, in particular, we will derive the elegant results of Murray Kemp and Ronald Jones by the application of the general optimum tariff formula. The optimum tax-tariff structure thus emerges solely from a consideration of the necessary marginal efficiency conditions at a Pareto optimum, independent of any assumptions regarding the existence of a community indifference map or of the need to treat factors of production in a special manner. We will focus primarily on their principal results for the case of non-specialization: the cases of specialization can be handled in a similar manner.

Both authors work within the Heckscher-Ohlin two-country, two-good, neoclassical trade model, but allow for the possibility of international differences in technology. Strictly speaking, they do not consider the problem of borrowing or lending over time but rather assume that the pattern of trade and, hence, of prices is the same in all periods As a result, at all times the value of goods imports plus rental payments from capital abroad is zero, i.e.:

$$(10)\qquad F \equiv Z_1 + pZ_2 + rZ_0 = 0,$$

where Z_1 and Z_2 are the flow of imports of the two consumption goods, Z_0 is the initial stock import of the capital good, p is the relative price of good two in terms of good one, and r is the rental flow on imported capital in terms of good one.[7]

From the balance of trade, (10), only two of the three import terms can vary independently and to make our results comparable with previous work we will eliminate the numeraire, Z_1, and consider p and r to be functions of Z_2 and Z_0 alone. Without loss of generality, a_1 is now identically zero and from equation (2), the optimal tax rate, τ, is given by:

$$(11)\qquad \tau = -a_2/(1+a_2)$$

Similarly, if θ is defined as the optimal tax on foreign earnings (assuming for the moment that the home country exports capital) then:

$$(12)\qquad \theta = -a_0$$

From (3), Section I:

$$(13)\quad a_2 \equiv \frac{Z_2}{p}\frac{\partial p}{\partial Z_2} + \frac{Z_0}{p}\frac{\partial r}{\partial Z_2} = \frac{1}{\eta_2} + \frac{\epsilon_2}{\eta_2},$$

where ϵ_2 is the foreign elasticity of import production with respect to borrowed capital, $-Z_0$, and η_2 is the total price elasticity of foreign import demand. In deriving (13) we have relied upon the assumption that the foreign country is non-specialized, implying that rentals are a function only of relative commodity prices and that $dr/dp = -\partial X_2/\partial Z_0$ where X_2 is the foreign production of the second good.[8] Combining (11) and (13) the optimal tariff is given by:

$$(14)\qquad \tau = -(1+\epsilon_2)/(1+\eta_2+\epsilon_2)$$

To derive the optimal tax on rentals we first note that holding Z_2 constant implies that:

$$(15)\qquad \frac{dZ_2}{dZ_0} = \frac{\partial Z_2}{\partial p}\frac{\partial p}{\partial Z_0} + \frac{\partial Z_2}{\partial Z_0} = 0,$$

and hence:

$$(16)\qquad \frac{\partial p}{\partial Z_0} = \frac{p}{Z_0}\frac{\epsilon_2}{\eta_2}$$

Thus, from (12) and (16):

$$(17)\qquad \begin{aligned}\theta = -a_0 &= -\left\{\frac{Z_2}{r}\frac{p}{Z_0}\frac{\epsilon_2}{\eta_2} + \frac{Z_0}{r}\frac{\partial X_2}{\partial Z_0}\frac{p}{Z_0}\frac{\epsilon_2}{\eta_2}\right\}\\ &= -\frac{1}{\mu_2}(1+\epsilon_2)\frac{\epsilon_2}{\eta_2},\end{aligned}$$

[7] The notation is the same as that of Section I, but since the model is in a stationary state, the time superscripts are superfluous.

[8] See Kemp, p. 807.

where $\mu_2 \equiv rZ_0/pZ_2$ is the ratio of the rental earnings on capital invested abroad to the value of exports. If the country is a debtor it is the earnings on foreign capital that must be taxed so that $1/(1-\theta)=r(1+a_0)$ which implies that:

$$(18)\quad \theta = \epsilon_2(1+\epsilon_2)/(\mu_2\eta_2 + \epsilon_2(1+\epsilon_2))$$

Equations (14), (17), and (18) are identical to those of Kemp.

In summary, approaching the problem of the optimal taxation of capital movements as a branch of optimal tariff theory has yielded the same results as those obtained in earlier work. Furthermore, adhering to the general point of view that the theory of optimal taxation of capital movements can be treated as a branch of optimal tariff leads one to the intuitive interpretation that the optimal pattern of taxes is determined directly from the relations of substitutability and complementarity embodied in the foreign elasticities of supply and demand. If, for example, the home country exports both capital and good 2 and if in the foreign country good 2 is relatively capital intensive, then good 2 and capital are substitutes, i.e., the cross elasticity, ϵ_2, is positive. Consequently, a tariff on good 2 necessitates the imposition of a tax on foreign earnings (e.g., on capital exports) to improve the terms of trade. It is instructive in this case to think of Robert Mundell's argument where, given non-specialization, a tax on commodity trade would be foiled by capital movements since capital and the capital intensive good are substitutes. If good 2 is relatively labor intensive abroad the analysis is slightly more complicated. Now domestic exports, good 2, and capital are complements in the foreign country and a tariff on one must be accompanied by a subsidy on the other, depending upon their relative importance in the balance of trade, and on the strength of what Jones has termed the "magnification effect."[9] If the home country is a capital importer, the analysis may be carried out by switching the terms "capital intensive" and "labor intensive," and "substitutes" and "complements" in the argument given above.

[9] See Jones (pp. 12-15) for a detailed treatment.

REFERENCES

R. E. Baldwin, "Equilibrium in International Trade: A Diagrammatic Analysis," *Quart. J. Econ.*, Nov. 1948, *62*, 748–62.

———, "The Role of Capital-Goods Trade in the Theory of International Trade," *Amer. Econ. Rev.*, Sept. 1966, *56*, 841–48.

G. Debreu, *The Theory of Value,* New York 1959.

W. Fellner, et al., *Ten Studies in the Tradition of Irving Fisher*, New York 1967.

I. Fisher, *The Rate of Interest,* New York 1907.

———, *The Theory of Interest,* New York 1930, reprinted 1967.

J. de V. Graaff, "On Optimum Tariff Structures," *Rev. Econ. Stud.*, 1949, *17*, 47–59.

———, *Theoretical Welfare Economics,* Cambridge 1957.

H. G. Johnson, "Optimum Tariffs and Retaliation," *Rev. Econ. Stud.*, 1954, *21*, 142–53.

R. Jones, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–38.

M. Kemp, *The Pure Theory of International Trade,* New Jersey 1964.

———, "The Gain from International Trade and Investment: A Neo-Heckscher-Ohlin Approach," *Amer. Econ. Rev.*, Sept. 1966, *56*, 788–809.

T. Koopmans, *Three Essays on the State of Economic Science,* New York 1957.

W. Leontief, "Theoretical Note on Time-Preference, Productivity of Capital, Stagnation, and Economic Growth," *Amer. Econ. Rev.*, Mar. 1958, *48*, 105–11.

G. D. A. MacDougall, "The Benefits and Costs of Private Investment from Abroad: A Theoretical Approach," *Econ. Rec.*, Mar. 1960, *36*, 13–35.

R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.*, June 1957, *47*, 321–35.

T. Negishi, "Foreign Investment and the Long-Run National Advantage," *Econ. Rec.*, Dec. 1965, *41*, 628–32.

P. A. Samuelson, "Irving Fisher and the Theory of Capital," in Fellner, et al., ed., *Ten Studies in the Tradition of Irving Fisher,* New York 1967.

R. M. Solow, *Capital Theory and the Rate of Return,* Rotterdam, 1964.

Public Utility Pricing and Output Under Risk: Comment

By Ralph Turvey*

In their paper in the March issue of this *Review*, Gardner Brown, Jr. and M. Bruce Johnson argue that "... the optimal price will always be lower and, with linear demand, the optimal output will generally be higher than their counterparts in the riskless model of traditional theory." (Their analysis relates to a single product produced at a constant unit operating cost of b and with a constant unit capacity cost of β. Thus in the riskless model, maximum welfare requires a price of $b+\beta$.)

Their result concerning price is an odd one. I accept that it follows from their assumptions but question the usefulness of these assumptions. The first step is to provide an intuitive understanding of their result. To do this, let us take the simplest possible case of risk where the demand curve has a probability of one half of being low and one half of being high. The price and capacity chosen must obviously be such that capacity falls between low demand and high demand at that price. If demand turns out to be low, consumers' surplus is maximized by producing up to the point where demand price equals b, since there is spare capacity and since capacity costs are fixed and hence irrelevant. If, on the other hand, demand turns out to be high, output will be limited by capacity, the excess demand being eliminated by rationing, so that price is irrelevant. Hence price can be determined so as to maximize consumers' surplus in the eventuality of low demand.

Capacity, on the other hand, being excessive when there is low demand, can be determined so as to maximize consumers' surplus in the eventuality of high demand. It should be increased up to the point where the expected gain in consumers' surplus from a unit increase in capacity $\frac{1}{2}(p-b)$ just equals the cost of that extra capacity β, where p is the (high) demand price at the level of capacity chosen.

I hope that this adequately conveys the essence of Brown and Johnson's argument. It rests upon the explicit and reasonable assumption that price has to be fixed before it is known whether demand is going to be high or low. But it also rests upon the implicit assumption that rationing is always preferable to price as a means of restraining consumption at times of high demand.

This assumption is not generally correct. A price greater than b will involve a loss of consumers' surplus at times of low demand but will diminish the amount of excess demand and hence the severity of rationing at times of high demand. There is thus a trade-off between the sacrifice of consumers' surplus on the one hand and the stringency of rationing on the other hand. The terms of this trade-off and consumers' attitudes to it will depend on the circumstances of the case, so they can be usefully discussed only in terms of particular real cases. Rationing in the case of electricity means power cuts; in the case of gas it may mean a pressure drop which entails the danger of explosions; in the case of telephone service it means that some calls cannot be made. In all three of these cases, tariffs are usually set high enough to keep down the risk of failure to a very low level, i.e., they are at a level considerably in excess of operating costs.[1] This does not impugn the principle set out above for determining capacity,[2] but it does suggest that Brown and Johnson's suggested

* The author is Joint Deputy Chairman of the National Board for Prices and Incomes in the United Kingdom.

[1] I have discussed the case of electricity in "Peak-Load Pricing" (p. 105) and in my book.

[2] But note that p is then to be interpreted as what customers will pay to avoid an unexpected failure of supply at short notice which may much exceed demand price in the ordinary sense.

pricing principle is not generally acceptable. As for their idea about covering capacity costs via transferable future rights in telephone calls or therms of gas, it was, I assume, meant to be funny.

REFERENCES

G. Brown, Jr. and M. B. Johnson, "Public Utility Pricing and Output under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

R. Turvey, "Peak-Load Pricing," *J. Polit. Econ.*, Jan./Feb. 1968, *76*, 101–13.

———, *Optimal Pricing and Investment in Electricity Supply*, London 1968, Cambridge, Mass. 1969.

Public Utility Pricing and Output Under Risk: Comment

By DAVID S. SALKEVER*

In their recent work on public utility pricing in this *Review*, Gardner Brown, Jr. and M. Bruce Johnson failed to point out one surprising implication of their model—namely, that peak-load pricing is inefficient under uncertainty. To see this, we simply postulate the existence of peak and off-peak stochastic demand functions, denoted by $X(P)+u$ and $Y(P)+v$, respectively, where u and v are random disturbances with zero mean and finite variance.[1] Since $X(P)+u$ is the peak demand, we assume $X(P)>Y(P)$ for all P satisfying $0\leq P<P^*$, where $X(P^*)=0$. The period of analysis is divided into peak and off-peak periods which take up fractions of the total period equal to e_x and $(1-e_x)$, respectively. Finally, let P_x be the peak price while P_y is the off-peak price.

The public utility seeks to maximize the weighted sum of expected consumers' surplus and expected net revenue during peak and off-peak periods, with the weights being e_x and $(1-e_x)$. We retain the assumption of constant marginal operating cost $(=b)$ and constant marginal capacity cost.

Where W is the weighted sum of expectations described above, first-order conditions for a maximum of W with respect to P_x and P_y are:

$$\text{(1)}\quad \frac{\partial W}{\partial P_x} = e_x\{P_xX'(P_x)F[Z-X(P_x)] - bX'(P_x)F[Z-X(P_x)]\} = 0;$$

$$\text{(2)}\quad \frac{\partial W}{\partial P_y} = (1-e_x)\{P_yY'(P_y)G[Z-Y(P_y)] - bY'(P_y)G[Z-Y(P_y)]\} = 0;$$

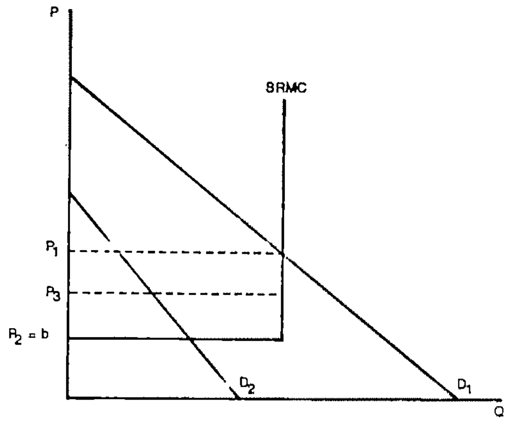


FIGURE 1

where Z is capacity and $F(u)$ and $G(v)$ are the cumulative densities of u and v.[2]

Solution of (1) and (2) for P_x and P_y yields $P_x=P_y=b$. Thus optimal peak and off-peak prices are equal. How can this result be reconciled with Oliver Williamson's conclusions that peak price exceeds off-peak price and that peak price exceeds marginal operating cost?

As I shall demonstrate, this difference in results is due to Williamson's tacit assumption that the market for the output of the public utility must be cleared at all times. When this assumption is dropped, even in the absence of uncertainty, consumers' surplus plus net revenue is maximized when both peak and off-peak price are equal to b. The introduction of uncertainty, however, implies that this is the only optimal pricing rule; with no uncertainty, there are an infinite number of optimal pricing rules.

Figure 1 demonstrates that the market clearing assumption may be dropped in the

* The author is a teaching fellow at Harvard University.

[1] We shall only discuss the case of additive uncertainty since the results are identical for the case of multiplicative uncertainty.

[2] These conditions are identical, except for the weights e_x and $(1-e_x)$, to Brown and Johnson's equation (16). For the derivation, see pp. 121–24.

peak period without any welfare loss. Where *SRMC* is the short-run marginal cost curve and the peak and off-peak demand curves are D_1 and D_2, Williamson asserts that P_1 and P_2 $(=b)$ are the optimal prices.[3] But it is clear that consumers' surplus plus net revenue is unchanged when peak price is reduced from P_1 to P_3.[4] In fact, this sum is unchanged for any $P \leq P_1$, since no greater output is produced.

Of course, each peak load price implies a particular distribution of welfare between consumers' surplus and net revenue. But in the absence of distributional preferences, any $P \leq P_1$ is an optimal peak-load price.[5]

A simple intuitive explanation now emerges for the Brown-Johnson result that price should always equal marginal operating cost when demand is random. For any random demand function, observed demand will intersect the *SRMC* curve on either its horizontal or vertical portions, depending upon the size of the disturbance term. If the price charged during the period when this random demand function prevails is greater than marginal operating cost, welfare (i.e. consumers' surplus plus net revenue) is not maximized unless observed demand intersects the vertical portion of *SRMC* at a sufficiently high level. Similarly, a price less than marginal operating cost is optimal only if the intersection does not occur on the horizontal portion of *SRMC*. But if price equals marginal operating cost, welfare is maximized regardless of where the intersection occurs. Since the point of intersection is not known in advance, price should always be set equal to marginal operating cost.

In conclusion, when demand is non-random, peak-load pricing may be justified either on distributional grounds or because it is cheaper to set a market-clearing price than to administer a rationing system that maximizes welfare. But when demand is random, peak-load pricing is simply inefficient. These conclusions are sensitive, however, to the cost specifications of the model. If marginal operating cost is not constant, different prices may be optimal during peak and off-peak periods when demand is random.[6] But if Williamson's cost assumptions are correct, a welfare justification of peak-load pricing (in the absence of distributional preferences) seems to depend on the existence of externalities caused by peak demands on the system. The welfare argument for "congestion taxes" on highways is a case in point.

[3] This follows from his assertion that "Optimal price in every sub-period is given by the intersection of the *SRMC* curve and the sub-period demand curve" (p. 821).

[4] We retain Brown and Johnson's assumption that when the market is not cleared, the product is rationed according to willingness to pay.

[5] A price greater than P_1 reduces output and thus reduces the area between *SRMC* and the demand curve, which happens to be consumers' surplus plus net revenue. (We may ignore capacity cost here since it is constant and independent of price.)

[6] When demand is non-random, welfare will still be maximized by setting peak and off-peak prices equal to marginal operating cost in the off-peak period, even if marginal operating cost is not constant.

REFERENCES

G. Brown, Jr. and M. B. Johnson, "Public Utility Pricing and Output Under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

Oliver E. Williamson, "Peak-Load Pricing and Optimal Capacity Under Indivisibility Constraints," *Amer. Econ. Rev.*, Sept. 1966, *56*, 810–27.

Public Utility Pricing and Output Under Risk: Reply

By M. BRUCE JOHNSON AND GARDNER BROWN, JR.

The purpose of this reply is to clarify the origins and implications of the pricing principle suggested in our earlier article in this *Review* and to argue that a modification of the model along the lines initiated by David Salkever will in some modest measure mitigate Ralph Turvey's complaint. We assumed that the presence of uncertainty requires the public utility to set both price P and capacity output Z before the position of the demand function is known. By hypothesis, therefore, capacity costs are sunk costs—independent of price. This assumption alone is responsible for the conclusion that optimal price P^* should equal short-run marginal cost b. These results are reminiscent of those of the traditional bridge problem: the capacity (investment) decision is made on the basis of the consumers' marginal valuation of a unit of capacity under the assumption that the long-run marginal cost will be equated to price; but *ex post*, capacity costs are ignored and price is set equal to *SRMC*. Reminiscent but not a replica since if we remove uncertainty from our model, the expected welfare function reduces to:

$$W = \int_{P}^{X^{-1}(0)} X(P)dP + PX(P) - bX(P) - \beta X(P)$$

where capacity costs are written $\beta X(P)$ rather than βZ because, under these circumstances, the public utility can set $Z = X(P)$ with confidence that excess demand or supply is impossible. Since optimal price in the certainty model is $P^* = b + \beta$, our pricing policy $P^* = b$ stems from the presence of rather than the amount of uncertainty as measured, say, by the range of the disturbance term.

However, there is a critical amount of uncertainty implicit in the model: our expected welfare function was based on the assumption that for some, however small, interval of values of the disturbance term the demand curves intersect the horizontal portion of the *SRMC* curve. This implies excess capacity at $P^* = b$ and represents the familiar off-peak model[1] of the peak-load pricing literature. Indeed, our model is a generalization[2] of Williamson's off-peak model to the case where the weights are interpreted as probabilities and the weighting function is held to be continuous.

Consider the following modification: if for given values of b and β the range of the disturbance term is so small that every demand curve $X(P)+U$ intersects the vertical portion of the *SRMC* curve, each *truncated* expectation in the expected welfare function becomes a *complete* expectation. The first-order conditions for maximization of the welfare function simplify accordingly: (1) price will equal that price just sufficient to clear the market when demand assumes its lowest position.[3] Thus $b < P^* < b + \beta$. (2) With a guarantee of no excess capacity, operating costs are known with certainty to be bZ; consequently, capacity is chosen to satisfy the equation: $X^{-1}(Z) = b + \beta$ or alternatively, capacity equals expected sales at the riskless price, $Z = X(b+\beta)$. This modification of the basic model obviously yields the stochastic counterpart to the firm-peak pricing problem.[4]

It is not surprising that Salkever finds peak and off-peak prices equal in the context of our original formulation. Unless the do-

[1] To be specific, by off-peak model we mean the case where the plant is not used to capacity when the off-peak demand function is in force. (See Williamson, p. 820.)

[2] The generalization must be qualified subject to the ways we allowed the disturbance term to enter the analysis as shown on p. 121 of our article.

[3] Let U_m be the smallest negative value (largest absolute value) that the disturbance term can take on in $X(P)+U$ and still have demand intersect *SRMC* on its vertical segment. Set capacity equal to this value of demand and solve for the price that satisfies the equation. Then $Z = X(P) + U_m \rightarrow P^* = X^{-1}(Z - U_m)$.

[4] In the firm-peak case plant is used to capacity whatever the position of the demand curve, see Williamson, p. 820.

mains of the density functions of the two random variables are sufficiently restricted, the distributions of peak demand $X(P_x)+U$ and off-peak demand $Y(P_y)+V$ will "overlap" so that knowledge of the weights e_x and $1-e_x$ provides little if any useful information to the public authority. This case is operationally indistinguishable from our original model as Salkever's results show. In order to make peak and off-peak stochastic demand functions nontrivial, it is necessary to further postulate that $X(P_x)+U$ invariably cuts the vertical segment of the *SRMC* curve (the firm-peak case) while $Y(P_y)+V$ cuts the horizontal portion of *SRMC* for all values of the disturbance term (the pure off-peak case). Redefining the problem in this fashion unquestionably results in a more plausible representation of peak and off-peak phenomena. Drawing on the equilibrium conditions of the original and modified models above this interpretation generates: (1) off-peak price equal to $SRMC,(P_y=b)$, (2) peak price equal to that price which clears the market when demand is at its lowest (peak) position, $(P_x=X^{-1}(Z-U_m)>b)$, and (3) capacity is chosen so that: $X^{-1}(Z)=b+\beta/e_x$. Capacity depends only[5] on peak demand and its weight; without uncertainty[6] peak price is greater: $P=b+\beta/e_x$. We conclude that in the stochastic peak-load pricing model developed here, although peak price is lower, capacity and off-peak price are the same as their counterparts in the model with certainty.

The conditional resurrection of multiple prices (i.e., $P_x>P_y=b$) to some extent meets Turvey's contention that price is generally preferable to rationing as a means of allocating available capacity among potential users when excess demand occurs. But peak price was chosen to clear the market for the lowest position of the demand curve and, consequently, must imply excess demand for all other positions. If the rules of the game are relaxed to allow the public authority to renegotiate price after the position of the demand function is known, naturally the effect of uncertainty is nullified and markets can and will be cleared. Furthermore, the utility could expect—with constant costs—to break even as the number of periods or trials increases. Capacity will still be chosen by our principle; the only change under this recontracting scheme is that if excess demand exists at the announced price, a surcharge must be added to eliminate the excess demand of that particular period. But it is well to remember that the final price can be quoted only after the random term is known—at the last minute, so to speak. This pricing scheme is of doubtful practical value; imagine announcing to the millions dialing their dearest relatives following a natural disaster that the available telephone circuits will be auctioned off to the highest bidders. Or that on the coldest night of the year the price of gas and electricity will be raised to an all-time high!

Obviously this is not the variety of market clearing arrangement that Turvey has in mind when he argues that price is generally preferable to nonprice rationing. We ignored the costs of rationing because they defied generalization and agree that the existence of such costs will increase price. However, this does not mean that excess demand will be elimiated as Turvey seems to suggest when he says: ". . . tariffs are usually set high enough to keep down the risk of failure to a very low level . . ." To price so high that there is rarely (never?) excess demand implies persistent excess supply (idle capacity) and an *ad hoc* decision by the public utility to remove the risk of failure. We might have more confidence in such a policy if it was derived from an optimizing model.

[5] If we allow off-peak demand to cut the vertical as well as the horizontal portion of the *SRMC* curve, the first-order condition pertaining to Z will include the truncated integral $\int_{Z-Y(b)}^{\infty} g(V)[Y^{-1}(Z-V)-b]dV$ indicating that capacity depends on both peak and off-peak demand.

[6] See Williamson, p. 821.

REFERENCES

G. Brown, Jr. and M. B. Johnson, "Public Utility Pricing and Output Under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

O. E. Williamson, "Peak-load Pricing and Optimal Capacity under Indivisibility Constraints," *Amer. Econ. Rev.*, Sept. 1966, *56*, 810–27.

Imperfect Asset Elasticity and Portfolio Theory

By MICHAEL A. KLEIN*

Although portfolio theory is a well-established subfield of the theory of decision making, it is a contention of this paper that the implications of the theory are not well suited to the analysis of a wide variety of applied problems. It will be argued here that the implicit assumption that all assets confronting an investor are in perfectly elastic supply to that investor constitutes a major weakness in the approach and is responsible for the limited applicability of the approach to applied research.

As an example, let us examine the implications of the mean, variance approach to portfolio theory[1] for cross-section differences in bank asset allocation choices. Assume the banker has a preference ordering over P, the rate of return on his portfolio, which can be represented by a utility function quadratic in P. Further assume that the asset universe confronting the banker consists of cash and a homogeneous private security (loan), whose risk and expected return are, in accordance with traditional theory, assumed exogenous to the bank. The banker must then choose the appropriate value for the loan/asset ratio, given his coefficient of risk aversion and the risk, return parameters of the private security.

It follows that cross-section differences in individual bank loan/asset ratios can be explained by cross-section differences in banker aversion towards risk and in differences in the risk and return characteristics of private securities. There can be little doubt that aversion to risk does differ among bankers but empirical verification of such a hypothesis is a highly formidable task. The second factor appears to be more easily testable. The problem, however, is that the risk and return characteristics of private securities are not truly exogenous to the bank. They are endogenous variables determined jointly by decision-making internal to the bank and certain characteristics of the economic environment which truly are exogenous to the individual bank.

This paper is exploratory. It attempts to provide a framework for the inclusion of imperfect asset elasticities within a general model of portfolio selection. This is set out in the two-asset case in the next section. Section II examines the implications of the theory for the 'dominant asset' theorem of James Tobin (pp. 196–200). Finally, Section III examines the theory's applicability to current problems in banking research.

I. *A Model of Imperfect Asset Supply*[2]

We assume that the bank faces a demand curve for loans which is a function of the contract rate of interest, r, and a vector of exogenous variables which influence the state of loan demand confronted by a particular bank. The scale and composition of the bank's liabilities are assumed given in the short run and it is further stipulated that the bank cannot discriminate among borrowers. That is, the model attempts to explain the choice between private securities and cash, not the choice among competing borrowers. We further assume for algebraic simplicity that default risk on private securities is identical for all securities and given exogenously to the bank.

Given these conditions, the bank must choose the loan/asset ratio, denoted by A, in

* Financial economist, Federal Deposit Insurance Corporation, on leave from Indiana University. The author would like to thank the Corporation for its generous research support and the referee for many helpful suggestions. Any remaining errors are the author's alone.

[1] The two major references are Harry Markowitz and James Tobin. The indebtedness of the present author to Tobin's paper is obvious.

[2] The model presented in this section is not intended to be interpreted as a theory of the banking firm. Such a theory is beyond the scope of this paper. Rather, the model attempts to illustrate the manner in which imperfect asset elasticities can be subsumed within the framework of a more general and ambitious model.

such a way as to maximize expected utility. The banker's utility function is given by

$$U(P) = (1 + b)P + bP^2 \tag{1}$$

where P is the rate of return on the portfolio and b is a coefficient of risk aversion which must lie between zero and minus unity for a risk averter.

The demand conditions facing the bank are summarized by equation (2) which

$$r = f(A), \quad f'(A) < 0 \tag{2}$$

is the analogue of a demand curve for loans, with the variable r denoting the contract rate of interest. Unless the banker assigns a zero probability to the events of partial or complete default on the private security, the expected return on loans, E_L, must be less than the contract rate of interest since the latter represents the *maximum* return the banker will receive. Generally,

$$E_L < r \quad \text{if } \sigma_L > 0 \tag{3}$$

where σ_L is a measure of default risk, the standard deviation of the probability distribution of loan payments. We earlier assumed that

$$\sigma_L = \bar{\sigma}_L \tag{4}$$

It follows from equation (2) and the above discussion that

$$E_L = h(A), \quad h'(A) < 0 \tag{5}$$

The bank wishes to choose A so that expected utility is maximized subject to certain constraints. Taking expected values on both sides of equation (1) we get

$$E[U(P)] = (1 + b)E_P + b[\sigma_P^2 + E_P^2] \tag{6}$$

where E_P and σ_P^2 represent the expected return and risk on the total portfolio. Equation (6) is to be maximized subject to

$$\sigma_P^2 = A^2\bar{\sigma}_L^2 \tag{7}$$

$$E_P = Ah(A) \tag{8}$$

Substituting (7) and (8) into (6) and differentiating the latter equation with respect to the decision variable, A, yields

$$\frac{dE[U(P)]}{dA} = [1+b][Ah'(A)+h(A)] + b[2A\sigma_L^2+2Ah(A)(Ah'(A)+h(A))] \tag{9}$$

The optimum A is found by setting (9) equal to zero and solving for A.

It will be easiest to illustrate if we assume that $h(A)$ is a linear function of A such as

$$h(A) = \alpha - \beta A \quad \beta > 0 \tag{10}$$

Substituting (10) into (9), collecting terms by order of A, and setting the result equal to zero yields

$$\frac{dE[U(P)]}{dA} = 4b\beta^2A^3 - 6b\alpha\beta A^2 + 2[b\sigma_L^2 + b\alpha^2 - \beta - b\beta]A + \alpha(1+b) = 0 \tag{11}$$

The second-order condition for a maximum follows:

$$\frac{d^2E[U(P)]}{dA^2} = 12b\beta^2A^2 - 12b\alpha\beta A + 2[b\sigma_L^2 + b\alpha^2 - \beta - b\beta] < 0 \tag{12}$$

If we divide (11) by $4b\beta^2$ we get

$$A^3 - \frac{3}{2}\frac{\alpha}{\beta}A^2 + \left[\frac{\sigma_L^2 b + \alpha^2 b - \beta - b\beta}{2b\beta^2}\right]A + \frac{\alpha(1 + b)}{4b\beta^2} = 0 \tag{13}$$

The solution is of the form

$$A = K(\alpha, b, \beta, \bar{\sigma}_L^2) \tag{14}$$

What is particularly interesting about this framework is that it allows explicitly for differences in loan demand confronting banks and for differences in the elasticity of demand for bank loans. As we shall see in Section III, this is of immediate applicability to problems in applied banking research.

We can determine qualitative comparative statics results without explicitly solving for

A. This can be done by differentiating the left-hand side of (13) with respect to both A and the particular variable we are interested in, and solving. We shall denote partial derivatives of the left-hand side of (13) by the symbol f with subscripts indicating the appriate variable. The following results are straightforward:[3]

$$(15.1)\qquad f_A = \frac{d^2E[U(P)]}{dA^2}\cdot\frac{1}{4b\beta^2} > 0$$

$$(15.2)\qquad f\sigma_L^2 = \frac{A}{2\beta^2} > 0$$

$$(15.3)\qquad f_b = \frac{4\beta^2[2\beta A - \alpha]}{16\beta^4 b^2} < 0$$

The latter result follows because $2\beta A - \alpha$ is equal to $-dE_P/dA$. If we impose the restriction that over the entire range of A, an increase in A increases the expected return on the portfolio, then f_b is negative. From the above we get

$$(16.1)\qquad \frac{dA}{db} = \frac{-f_b}{f_A} > 0$$

$$(16.2)\qquad \frac{dA}{d\sigma_L^2} = \frac{-f\sigma_L^2}{f_A} < 0$$

With respect to the demand parameters, an increase in α or a reduction in β represent an unambiguous improvement in the banker's opportunity set. That is, for every A, E_L would be greater than before. It follows that Tobin's comparative statics results are applicable[4] and that

$$(16.3)\qquad \frac{dA}{d\alpha} > 0$$

$$(16.4)\qquad \frac{dA}{d\beta} < 0$$

We shall return to the implications of the model for bank portfolio behavior in Section III.

[3] I am particularly indebted to the referee for simplifying the argument at this point.

[4] See Tobin, equation 3.16, p. 194.

II. *Implications of the Theory for the Dominant Asset Theorem*

The framework provided above has the unfortunate implication that one of the most powerful theorems of traditional portfolio theory is no longer valid when asset supplies are not perfectly elastic. We refer to the dominant asset theorem of Tobin. Basically, the theorem states that when there are multiple alternatives to cash, these can be treated as one asset which is, in fact, a weighted average of the alternatives ***and that the weights can be chosen without reference to the bank utility function***. The latter serves only to determine the proportion of cash in the portfolio.

To illustrate, imagine we have two assets, loans and government securities, and let E_L, σ_L^2, E_g, σ_g^2 denote the expected return and risk on loans and government securities respectively. Let X_L^P equal the proportion of loans in the portfolio and X_g^P the proportion of government securities. Assuming that cash is riskless and without return, the risk and return on the total portfolio can be written as[5]

$$(17)\qquad E_P = X_g^P E_g + X_L^P E_L$$

$$(18)\qquad \sigma_P^2 = (X_g^P)^2\sigma_g^2 + (X_L^P)^2\sigma_L^2$$

Letting A now represent the proportion of earning assets to total assets, we have

$$(19)\qquad X_g^P + X_L^P = A$$

If we minimize (18) for every E_P and solve for X_g^P and X_L^P, we get

$$(20)\qquad X_g^P = \frac{E_g\sigma_L^2}{E_L\sigma_g^2}\cdot X_L^P$$

We cannot yet determine either X_g^P or X_L^P since that would require knowledge of A and that requires that we introduce the utility function. Nevertheless, we have determined the ***relative*** mix of loans and government securities and can, therefore, treat the next decision as if it involved cash and just one risky asset.

[5] We assume for algebraic simplicity that the returns on loans and government securities are uncorrelated.

If one of the assets, say loans, has less than perfect elasticity then $E_L = h(A)$. It follows that the dichotomization of the portfolio choice problem is no longer valid. That is, E_L cannot be determined without first knowing A, and this requires the introduction of the bank utility function.

III. *Implications of the Theory for Banking Research*

The approach presented above is more productive in terms of theory and in its implications for empirical work than is the traditional one. With respect to the former, it is more satisfying theoretically because it relates the bank decision variable to variables which are truly exogenous to the bank. In this context it is important to understand that equation (14) is not to be construed as a loan supply curve. It is, rather, a reduced form equation relating the equilibrium loan/asset ratio to variables which are exogenous to both the individual bank and its borrowers. Given A, the contract rate of interest can then be derived through equation (2).

The above approach is also more productive in terms of empirical implications than is the traditional approach. For example, there has been some controversy in the banking literature as to whether market concentration or differences in regional demand conditions can account for cross-section differences in bank loan rates.[6] Traditional theory simply does not yield an estimating equation which accounts for the influence of either of these variables. As a result, the empirical literature on the subject has proceeded divorced from economic theory. The approach presented here is capable of integrating the two hypotheses. Differences in loan demand would be reflected in different values of α, and differences in market concentration can be expected to affect the point elasticity of demand for loans and, therefore, the slope of the h function, or β.

Finally, it should be noted that although we have used a commercial bank to illustrate the problem, most financial intermediaries face an asset universe which consists, in part, of assets which are in imperfectly elastic supply to that intermediary. By specifically allowing for differences in asset supply across such intermediaries, we have developed a framework which should be more productive with respect to the generation of empirical hypotheses and, therefore, more useful in applied work.

[6] See, for example, Theodore Flechsig and Franklin Edwards.

REFERENCES

F. R. Edwards, *Concentration and Competition in Commercial Banking*, Res. Rep. No. 26, Federal Reserve Bank of Boston 1964.

T. G. Flechsig, "The Effect of Concentration on Bank Loan Rates," *J. Finance*, May 1965, *20*, 298–311.

H. Markowitz, *Portfolio Selection*, New York 1959.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," in Richard S. Thorne, ed., *Monetary Theory and Policy*, New York 1966, pp. 178–205.

Land in Fellner's Model of Economic Growth: Comment

By Roger A. McCain*

In a recent article in this *Review*, William Fellner explored the implications of the neoclassical theory of economic growth for the measurement of technological change. His model assumed two inputs, capital and labor, and constant returns to scale. Fellner speculated that capital and land are readily substitutable, and hence no separate rent category of income, or land category of inputs, need be provided. It is the purpose of this note to demonstrate that when land is introduced explicitly as a factor of production in a model with factor-augmenting technological change, it can be dealt with in a way that validates Fellner's model as regards the capital and labor interactions; but moreover the introduction of land as a separate factor of production contributes to the predictivity of the model by resolving, at least in part, a criticism of neoclassical growth theory made by B. Frey.

A Constant Elasticity of Substitution (CES) production function for three factors of production adequate to the purpose of this paper may be written

$$(1)\quad Y = [(E_N N)^\rho + (E_K K)^\rho + (E_L L)^\rho]^{1/\rho}, \quad \rho < 0,$$

where

N = labor
E_N = the efficiency of labor relative to some base period
K = capital
E_K = the efficiency of capital relative to the same base period
L = land
E_L = the efficiency of land relative to some base period
Y = the national income

This production function has the property that the elasticity of substitution between any pair of factors, given any quantity of the third factor, is $1/1-\rho$.[1]

The other elements of the model are as follows:

w = the real wage
r = the rate of interest
n = real rent
$e_N = (E_N N/Y)(\partial Y/\partial E_N N) \equiv$ the elasticity of output with respect to augmented labor
$e_K = (E_K K/Y)(\partial Y/\partial E_K K) \equiv$ the same for augmented capital
$e_L = (E_L L/Y)(\partial Y/\partial E_L L) \equiv$ the same for augmented land

G_X, (X any variable), represents the proportionate growth rate of X.

It is assumed that returns to scale are constant so that $e_K + e_N + e_L = 1$. It is assumed that max $G_{E_N} = g(G_{E_K}, G_{E_L})$, i.e., the innovation-possibility frontier applies to land as well as capital and labor. As usual the area under the innovation-possibility frontier is assumed to be convex.

Three behavioral assumptions are made. First is the usual assumption that capital formation is a constant proportion of national income and other national income is consumed; second that each factor is paid its marginal product. The third is the profit-maximizing assumption that innovations are so chosen that $\partial G_{E_K}/\partial G_{E_N} = -wE_N N/rE_K K$;

* The author is lecturer in economics at Western Washington State College. He wishes to express gratitude to an anonymous referee and to Michael Blundell, Michael Starr, and William Unger, formerly economics majors at the State University College at Geneseo, New York.

[1] The identity of the elasticities of substitution as between any pair of inputs may seem to some, as it seems to the author, uncomfortably restrictive. However, less restrictive generalizations of the two-factor *CES* function present some problems (see Robert M. Solow, pp. 42–45).

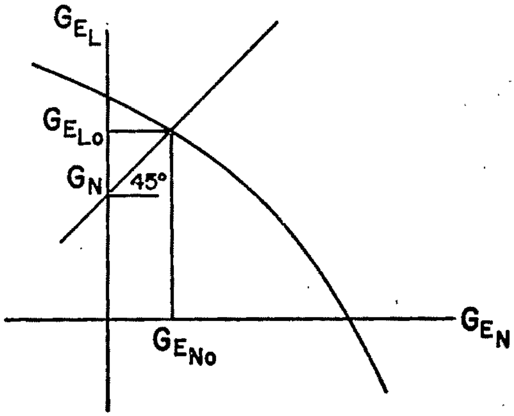


FIGURE 1. PARTIAL EQUILIBRIUM OF LABOR AND LAND AUGMENTATION

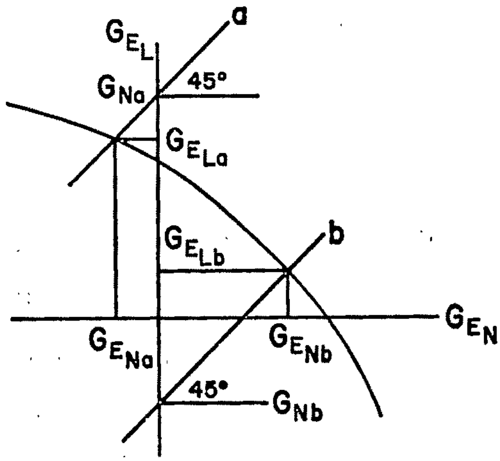


FIGURE 2. PARTIAL EQUILIBRIUM WITH VERY RAPID AND WITH NEGATIVE POPULATION GROWTH

$\partial G_{B_L}/\partial G_{B_N} = -wE_N N/nE_L L$. This is a routine generalization of Fellner's rule for two factors.

Now it is clear that

$$(2)\quad G_Y = e_K(G_K + G_{B_K}) + e_N(G_N + G_{B_N}) + e_L(G_{B_L}),$$

since the quantity of land cannot be increased. For the moment let us leave the first term alone and assume that the rate of capital augmentation is given at an arbitrary rate. Then the effective innovation-possibility frontier for labor and land would be a cross-section of the full three-factor innovation-possibility frontier (or innovation-possibility surface if you will). Let us investigate the partial equilibrium of land-augmentation and labor-augmentation with capital-augmentation at the given level.

In order that G_{B_N} and G_{B_L} be constant, it is necessary that $wE_N N/nE_L L$ be constant, but, with the elasticity of substitution less than unity, this requires that $G_{(B_N N/B_L L)} = 0$. That is,

$$(3)\qquad G_{B_L} = G_N + G_{B_N}$$

In case the rate of land augmentation is greater than the rate of population growth when the rate of labor-augmentation is zero, we will have a situation like that of Figure 1. At the intersection of the partial innovation-possibility frontier with the 45° line, $G_{B_L} = G_N + G_{B_N}$. There are other possibilities as in Figure 2: the growth rate of the labor input may be so great that there will be labor disaugmentation (a), or the growth rate of the labor input may be negative (b). The smaller is the growth rate of the labor force the more rapid will be labor-augmentation, which is, after all, just as one would suspect.

However, substituting $G_N + G_{B_N}$ for G_{B_L} in (2), we obtain

$$(4)\quad G_Y = e_K(G_K + G_{B_K}) + (e_N + e_L)(G_N + G_{B_N})$$

In "golden age" equilibrium, that is[2]

$$(5)\qquad G_Y = \frac{e_K}{1 - e_K} G_{B_K} + G_N + G_{B_N}$$

[2] Given the values of G_{B_K}, G_N, and G_{B_N}, $\frac{Y}{K}$ will tend to a level such that $G_K = G_Y$. This is the equilibrium value of $\frac{Y}{K}$, and is stable (see Fellner, pp. 1077, 1080). Thus we have

$$G_Y = e_K G_Y + e_K G_{B_K} + (e_N + e_L)(G_N + G_{B_N}),$$

that is,

$$(1 - e_K)G_Y = e_K G_{B_K} + (e_N + e_L)(G_N + G_{B_N}),$$

or

$$G_Y = \frac{e_K}{1 - e_K} G_{B_K} + \frac{e_N + e_L}{1 - e_K}(G_N + G_{B_N}).$$

from the constancy of returns to scale, $e_K + e_N + e_L = 1$, or $e_N + e_L = 1 - e_K$.

Arguments perfectly analogous to those of Fellner establish that G_{B_K} will tend to zero. They need be qualified only in that each decrease of G_{B_K} will shift outward the partial innovation-possibility frontier for land and labor, equally increasing the rates of augmentation of land and labor. Full equilibrium of factor-augmentation is shown in Figure 3; the final expression for the growth rate of national income is the familiar

$$G_Y = G_N + G_{B_N} \tag{6}$$

All of the usual results follow, including, in particular, the equating of the growth rate of income per worker and of the wage with G_{B_N}.

Thus the approach of Fellner is workable with a separate input of land; his assumption of constant returns to scale vis-à-vis labor and capital is unnecessary. Within a single economy, it may be difficult to determine whether in fact the interpretation of this paper or that of Fellner is the proper one. Equation (4) above suggests that an attempt to measure the relative importance of labor growth and capital growth in determining income growth by estimating their elasticities may lead to results which exaggerate the importance of labor growth.

However, the model presented here may lead to differences in the growth path of different countries. Frey has criticized neoclassical growth theories on the grounds that they do not explain differences in growth patterns between countries (p. 74). Every country would be expected to have an equal rate of growth of income per worker, except insofar as they experienced different rates of (otherwise unobservable) technological improvement. Additionally, in the model as given by Fellner, the distribution of income would depend only on the production-possibility frontier, so that international differences of income-distribution can be explained only by differences in the innovation-possibility frontier. This is perhaps a bit disturbing because the international movement of technical personnel and innovations is a common experience. It is at least an attractive hypothesis that technology is a pure public good, even at the international level.

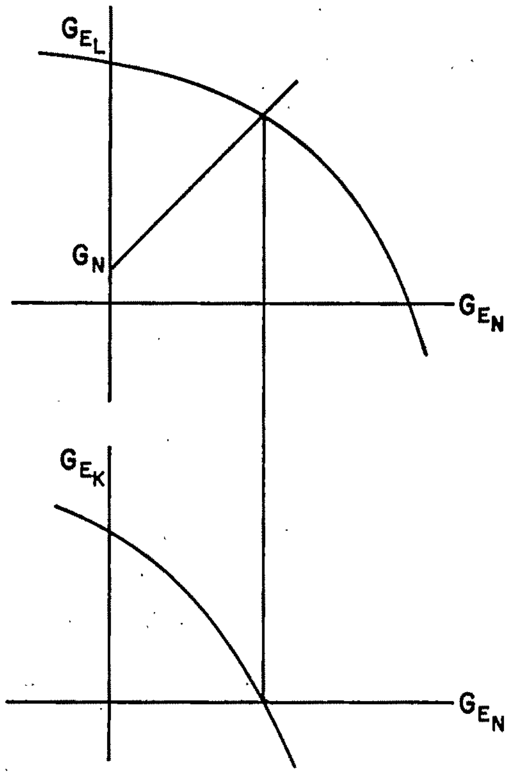


FIGURE 3. FULL EQUILIBRIUM OF AUGMENTATION OF THREE FACTORS

However, nations clearly differ in their demographic makeup, cultures and existing technological hardware; perhaps they differ adequately to assure substantially different innovation-possibility frontiers.

The model given in this note indicates a strong negative correlation of labor force growth with technological progress. The rate of growth overall will vary directly with the rate of population growth, however. Recall that G_{B_N} and thus G_Y are determinate whenever G_N is known. Assuming $G_{B_N} = \max G_{B_N}$ for given G_{B_K}, G_{B_L} (i.e. the economy is on its innovation-possibility frontier)

$$G_{B_N} = g(G_{B_K}, G_{B_L}) = g(0, G_N + G_{B_N})$$

$$\frac{\partial G_{BN}}{\partial G_N} = \frac{\partial g}{\partial G_{B_L}} \frac{\partial G_{BN}}{\partial G_N} + \frac{\partial g}{\partial G_{B_L}}$$

That is,

$$\frac{\partial G_{BN}}{\partial G_N} = \frac{\dfrac{\partial g}{\partial G_{BL}}}{1 - \dfrac{\partial g}{\partial G_{B_L}}}$$

Hence

$$\text{(6a)} \qquad \frac{\partial G_Y}{\partial G_N} = 1 + \frac{\dfrac{\partial g}{\partial G_{B_L}}}{1 - \dfrac{\partial g}{\partial G_{B_L}}}$$

Since $\partial g/\partial G_{BL}$ is negative, that is

$$\text{(6b)} \qquad \frac{\partial G_Y}{\partial G_N} = 1 - \frac{\left|\dfrac{\partial g}{\partial G_{B_L}}\right|}{1 + \left|\dfrac{\partial g}{\partial G_{B_L}}\right|}$$

As G_N increases, so does G_{BL}, and so $\partial g/\partial G_{BL}$, which is the inverse of the slope of the partial innovation-frontier will increase due to the convexity of the innovation-possibility frontier.[3] As G_N increases, then, G_Y will increase also, but at a slower rate.

In other terms, suppose that G_N should increase. Once a new golden-age equilibrium is achieved, G_{BN} must be diminished, and hence the rate of increase of per capita output (i.e., of the standard of living) must be diminished. If it did not, G_{BL} would have to increase for shares-equilibrium. However, G_{BK} is zero in both equilibria. Hence a constancy or increase in G_{BN} and increase of G_{BL} would be inconsistent with the convexity of the innovation-possibility set.

On the other hand, G_{BL} must increase. Decreasing G_{BN}, constant G_{BK}, and constant or decreasing G_{BL} would be consistent with a convex possibility set, but would be inconsistant with either: a positive share of rent in national output (since a vertical or upward-sloping frontier is implied); or the assumption that the economy remains on the frontier. Thus shares equilibrium requires a higher G_{BL} (and a higher G_N+G_{BN}, in spite of the lower G_{BN}).

These propositions, both of which appear to be of importance, cannot be derived from Fellner's model.

Additionally, the model may account for some differences in the distribution of income among countries with identical innovation-possibility frontiers. If one country has a higher rate of population growth than another country, it will occupy a different point on the innovation-possibility surface, and hence its distribution of income will differ. Moreover, the higher is the rate of population growth, the less is $\partial G_L/\partial G_N$, and hence the smaller is the share of labor relative to that of capital.

The symmetry of the land, labor, capital triad of classical economics invites further development. In fact, a model rather like the classical one can be based on the basic elements of this model. Instead of an exogenously determined rate of population growth, assume that the rate of population growth is a nondecreasing function of income per capita. Assume that the rate of labor force participation is constant, so that the labor input grows at the same rate as population. If, due to some disturbance of the system, income per capita were to rise high enough that G_N is greater than the level of G_{BL} associated with zero labor augmentation (point P in Figure 4), we would have a Ricardian-Malthusian case. Such a case is shown in Figure 2, *a*. Labor is being disaugmented, so wage rates and incomes per capita are falling. Thus the rate of population growth will be falling. Although the effective supply of land is growing at the same rate as the effective supply of labor, the effective supply of land is increasing less rapidly than is the population. Since the model treats land as homogenous, however, there is no scope of extension of the margin of cultivation within it. However, as population-growth slows, labor-augmentation increases, so $|\partial G_{BL}/\partial G_{BN}|$ increases and hence so does the share of wages relative to that of rent.

Conversely, if the initial conditions involved a per capita income low enough that G_N were less than P, the case would be that

[3] I am indebted to the referee for pointing out my error on this point on an earlier draft.

of Figure 1. There wages and product per capita are rising as labor is augmented. The rate of population growth will be rising. The only stable rate of population growth is P in Figure 4. Figure 4 then shows the full Malthusian equilibrium, which has the following characteristics:

(a) $G_{BK} = G_{BN} = 0$. All technological progress appears as augmentation of land.
(b) $G_{BL} = G_N$
(c) income per capita is constant at the corresponding level
(d) the distribution of incomes will be determined by the appropriate partial derivatives of g.

Of course, $G_N = 0$ and income per capita is at the "subsistence" level (i.e. that level at which cultural, hygienic and other considerations determine a zero rate of natural increase) if, and only if, $G_{BL} = 0$ when $G_{BN} = G_{BK} = 0$; i.e., if no technological progress occurs. This means, of course, that the innovation possibility frontier passes through the origin. That being true, though no technological *progress* is possibile, technological *change* is possible since one factor can be augmented at the cost of disaugmenting one or more others.

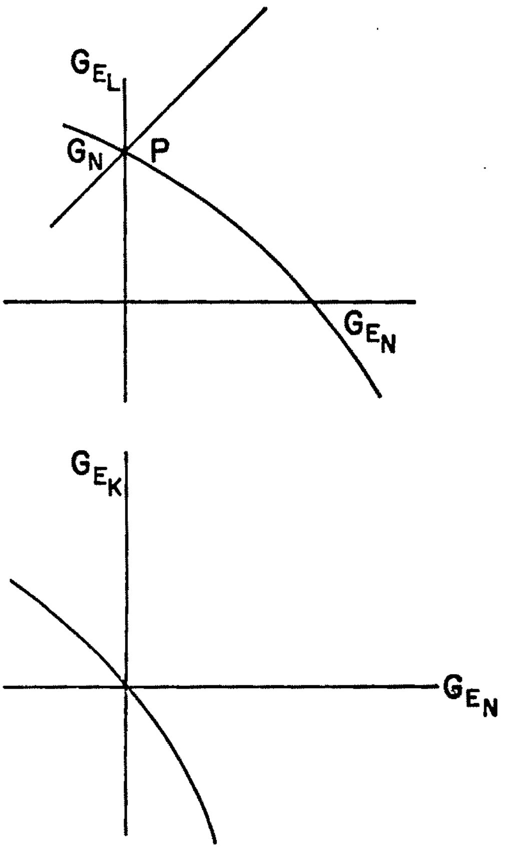


FIGURE 4. FULL MALTHUSIAN EQUILIBRIUM

In summary, a golden-age equilibrium in a model with land and factor augmentation is characterized by augmentation of land at a rate equal to the sum of the rate of growth of the labor input and the rate of augmentation of labor. The rate of capital-augmentation is zero. An increase in the rate of growth of the labor input dictates an increase in the rate of augmentation of land. From that follows a contraction of the partial innovation-possibility frontier for land and capital, and a decrease in the rate of labor augmentation and per capita growth. If the growth rate of the labor input is supposed to be an endogenous variable along classical lines, the rate of labor force augmentation tends to zero along with the rate of capital augmentation, as all innovations are concentrated on the improvement of the land. However, in that case the division of income among land, labor and capital is determined, not by the proportion of labor and land, but by the growth rate of labor.

REFERENCES

W. Fellner, "Technological Progress and Recent Growth Theories," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1073–98.

B. Frey, "Eine politische Theorie des wirtschaftlichen Wachstums," *Kyklos*, 1968, *21*, Fasc. 1, 70–101.

R. M. Solow, "Some Recent Developments in the Theory of Production," in M. Brown, ed., *The Theory and Empirical Analysis of Production*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 31, New York 1967.

Land in Fellner's Model of Economic Growth: Reply and Further Observations

By William Fellner*

Roger McCain's valuable contribution demonstrates that in the framework of the recent theory of induced innovations it is indeed possible to accommodate "land." He convincingly calls attention to some of the further results—expressed in part as testable propositions—which are obtained by his extension of shares-equilibrium models. I will merely add a reminder of a *modification* which the two-factor results require if they are to be carried over into a three-factor system.

I will make use of this opportunity also to correct a statement I made in the first complete paragraph of page 1094 of my 1967 article. Aside from this specific correction—i.e., in all other respects—the reader preferring the three-factor approach may not find it too difficult to transpose my analysis, and that of other authors, into the new model. McCain has provided much guidance for this transposition.

In the golden-age equilibrium emerging in my two-factor model, the proportionate rate of increase of total factor-productivity must equal the algebraic product of the proportionate rate of increase of output per unit of labor input with the elasticity of output with respect to labor. As I stated in my article, the reason for this is that the function which technological progress (as measured by the growth of total factor-productivity) performs in a two-factor model is that of "filling in" the gap between the growth-rate of physical capital (which equals that of output) on the one hand, and the lower growth rate of the labor input, on the other hand.

We should now take notice of the fact that in McCain's system this conclusion ceases to be valid because here technological progress (as measured by the growth of total factor-productivity) performs *also* the additional function of making up for the gap between the growth rate of physical capital (which equals the growth-rate of output) on the one hand, and the growth-rate of *land*, in natural units, on the other. Hence, the *ratio* of the rate of increase of total factor-productivity to the rate of increase of output per unit of labor here gets to be a *larger* fraction than the equivalent of the labor share (provided this share is an acceptable proxy for the elasticity of output with respect to labor). The McCain approach destroys some of the simplicity of the two-factor results, but this speaks for the approach, because the two-factor analysis has the rather disturbing implication that, in addition to the technological progress which is explicitly recognized, there is "in the background" just enough tacitly assumed progress to eliminate the consequences of the relative land-scarcity. In my article I did point out that this implication is a blemish which my analysis has in common with that of other authors. McCain's system has no such implication (is free from this blemish); it explicitly includes the land-augmenting effect of progress.

If my thinking had moved along McCain's lines, I would have been more clearly aware that there exist two reasons why for the period 1884–1957 taken as a whole, i.e., for the period I considered in the paragraph which I am now correcting, the ratio of total factor-productivity-growth to the growth of output-per-labor must exceed the labor-share. *First*, output-growth has tended to exceed even the growth of *reproducible* capital. By how much depends on somewhat controversial methodological decisions, but at any rate this upward deviating tendency from golden age suggests in itself that progress has performed a stronger function that that of making up the gap between the output-trend and the labor-input trend. This

* Yale University

point I did discuss in my article. But, *secondly*, as McCain shows, not even in golden age would the ratio of total factor-productivity growth to the growth of output-per-labor be down to the level of the labor-share, because output would still be rising in excess of land (in natural units), the difference being made up by the land-augmenting quality of progress.

Calculating the ratio in question for the period 1884–1957 with inclusion of John Kendrick's estimate of land, and generally using his data, the correct estimate of the ratio is 0.86.[1] This is not the figure I obtained, and it is an *appreciably* higher figure than any reasonable estimate of the labor-share for the period. The labor-share may not have been far from the low end of the 0.70 to 0.75 range.

Even if we did try to eliminate land, and if we tried to follow through the two-factor implication by which the land-augmenting aspect of progress is left in the background, the ratio would probably not decline below 0.79. I say "probably" because I have no way of undertaking that correction with any great precision. However, though now under McCain's influence I feel no great inclination to follow through these two-factor implications, I think I could do a somewhat better job of adjustment than I was able to do at the time of writing my article when I was thinking in those terms and proceeded as if two (from my point of view) "undesirable" properties of the data roughly cancelled. At that time I obtained (in the paragraph which I am now correcting) a ratio of 0.71, instead of what probably should have been "0.79 when land-augmentation remains in the background." Yet it seems to me that what now deserves emphasis is the fact that if along McCain's lines land-augmentation is made explicit, and if Kendrick's estimates are used, then the ratio comes out unequivocally at 0.86. The high level of this ratio is in good part a consequence of the fact that not even in golden age would the ratio decline to the level of labor's share in income.

[1] When land is included we may take the following ratio. Numerator: the annual compound rate of increase of Kendrick's total factor-productivity (which excludes from the growth of productivity the effect of allocational shifts between the sectors which he distinguishes). Denominator: the annual compound rate of increase of Kendrick's output per unit of labor input (which is so corrected as to include into the growth of input the movement from lower to higher productivity sectors). Numerator: 1.52 percent; denominator: 1.77 percent. See John W. Kendrick, Table A XXII, pp. 333–335. Alternatively we could take, with essentially the same result, the ratio of the two magnitudes defined in this footnote but so as to include into productivity-growth the effect of the intersector shifts and hence not raise the "labor input" when relative allocation changes in favor of higher-productivity sectors (see Kendrick's "Supplement" to his foregoing Table, p. 336).

REFERENCE

J. W. Kendrick, *Productivity Trends in the United States*, Princeton 1961.

Realized Interest Rates and Bondholders' Returns

By ROMAN L. WEIL*

Consider these two questions: What have real interest rates been throughout this century? What have been the real returns to bondholders?

As far as I know these questions have gone unanswered. This article reports my answers and describes the methods employed in their derivation. That these two are not the same question (even when the first is answered from bond data) but that they are near enough to be answered together will be made clearer below. Section I is devoted to the first question and Section II to the second.

I. *Realized Interest Rates*

I distinguish among ***instantaneous real***, ***anticipated real***, and ***actual*** or ***realized*** rates. To define these terms, first define the nominal rate of return (or interest) on an asset, denoted $i(t)$, by

$$i(\mathrm{t}) = \frac{P_A(\mathrm{t}) - P_A(\mathrm{t}-1)}{P_A(\mathrm{t}-1)} \tag{1}$$

or, expressed as a continuous rate over k periods,

$$i(\mathrm{t}) = \frac{1}{k}\, ln\, \frac{P_A(\mathrm{t})}{P_A(\mathrm{t}-k)} \tag{2}$$

when $P_A(\mathrm{t})$ is the price of the asset at time t.

If $p(\mathrm{t})$ is the price level at time t, the ***instantaneous real*** rate of interest $r(\mathrm{t})$ is defined for a given ***epoch*** t by

$$r(\mathrm{t}) = \left[i(\mathrm{t}) - \frac{1}{p}\,\frac{dp(\mathrm{t})}{d\mathrm{t}}\right] \div \left[1 + \frac{1}{p}\,\frac{dp(\mathrm{t})}{d\mathrm{t}}\right] \tag{3}$$

To measure it requires continuous data on nominal interest rates and price changes. The ***anticipated real*** rate is defined for a time ***span*** and is the nominal rate over the span less the rate of change of prices over the span divided by a term which accounts for appreciation or depreciation on interest payments. See the discussion by Irving Fisher on pages 41–44 and chapter 19 of his *The Theory of Interest*.

The realized rate can be measured but the definition above implies at least two ways to measure it. The ***anticipated*** real rate is the nominal rate for a span less the anticipated price rise over that span divided by the adjustment term. The ***actual*** or ***realized*** rate

* The author is associate professor of management and information science in the Graduate School of Business at the University of Chicago. He thanks Lawrence Fisher, David Green, Reuben Kessel, Arthur Laffer, Lester Lave, Charles Nelson, Joel Segall, David Shanno, and the referee for useful insights. Research support was provided by the National Science Foundation. These findings were presented to the Third Annual Advanced Management Seminar (Graduate School of Business, University of Chicago, March 1969).

After this article was submitted, John P. Gould read it and found an error in the original equation (3). I had omitted the division which corrects for appreciation or depreciation on the interest payments.

James C. Ellert also noted this error. In addition, Ellert persuaded me that for 1900–20 the Wholesale Price Index as shown in Series E-13 of the U.S. Bureau of the Census is more appropriate for ascertaining rates of price change than the *GNP* Deflator, Series F-5, which I used. E-13 is better than F-5 in the period 1900–20 because the numbers in F-5 are five-year averages whereas those in E-13 are annual. Ellert redid the calculations including the division term from equation (3) and Series E-13. The text is suitably changed but the top diagram in Figure 1 uses the *GNP* Deflator throughout. Compare the numbers in column (2) of Table 1 to the points in the top chart in Figure 1 to see the difference. The regularity shown in the *GNP* Deflator for the years 1900–15 is a consequence of the data being given in five-year intervals and my derivation of one-year data from them.

The results in Table 1, columns (1), (2), (3), (5), (6), and (9) are Ellert's calculations using my linking procedures as explained in the text. Ellert finds my inability to explain faulty anticipations of price-rate changes to be unsatisfying, and he is attempting to explain nominal rates with lagged price changes.

For obvious reasons, I appreciate Gould's and Ellert's efforts.

over a span would be the nominal rate less the actual change in prices over the span divided by the adjustment term. There is no a priori reason to suspect that the measured anticipated real and measured actual rates would be identical. Nevertheless, if over the long run one believes, as I do, that there are no permanent biases in anticipations, then the anticipations are fulfilled. If bondholders do not systematically over- or underestimate price changes in the long run, then the measured anticipated rate will be an unbiased estimate of the actual or realized rate. See the discussion by Frederick R. Macaulay in chapter 6 of his *Interest Rates, Bond Yields, and Stock Prices*.

Below, I measure anticipated and actual rates calculated from discrete rates of interest and price changes over one-year intervals. Continuous rates over k periods could just as well be calculated from the formula

$$r(\mathrm{t}) = \frac{1}{k} \ln \left\{ [P(\mathrm{t})/P(\mathrm{t}-k)] \cdot [L(\mathrm{t}-k)/L(\mathrm{t})] \right\} \tag{4}$$

where P is the asset price and L is the price level.[1]

Data and Results.

I have approximated $i(\mathrm{t})$ with David Durand's series on high grade corporate bond *one-year* yields to maturity as brought up to date and reported by the National Industrial Conference Board. These numbers are shown in column (4) of Table 1. There are many yield-to-maturity series for various bonds and for terms to maturity longer than one year but because the term structure of interest rates is generally not flat, accurate measures of $i(\mathrm{t})$ for one year cannot be made from them. That is, calculations of one-year rates from longer-than-one-year bonds are not good. The Durand data are subject to varying degrees of error. Durand (p. 14) points out that in the best years, the numbers shown are correct plus or minus $\frac{1}{4}$ percent (25 basis points) and in the worst may be off by as much as 1 percent (100 basis points). The Durand data are the best I could find for this purpose. Nevertheless, the reader should not be lulled by their published, "official" nature. Durand himself points out their shortcomings and the researcher who plans to use these data should, at a minimum, read Durand's caveats.

The price change data are troublesome. In Table 1 column (1) are shown levels of the *GNP* price deflator for each year since 1929.[2] For 1900–29 I have used Series E-13 of the U.S. Census Bureau's *Historical Statistics of the United States, Colonial Times to 1957* and *Continuation and Revisions*. Series E-13 is the Wholesale Price Index.

So that short-run price changes will be more apparent, I have calculated and show in column (2) the price levels of column (1) with the trend removed. The price level for 1900 is 29.01 and for 1968 is 120.5. Prices rose at an average annual rate of 2.09 percent compounded continuously. In column (2), the price levels for 1900 and 1968 are set to 100 and had prices grown uniformly at a continuous rate of 2.09 percent per year, all the numbers in column (2) would be 100.[3]

[1] Alternatively, the continuous rate can be calculated from the formula for the geometric mean of a continuous function. If the geometric mean be used for periods t_0 to t, then the mean rate is defined by

$$r(\mathrm{t}) = \exp \left\{ (1/(\mathrm{t}-\mathrm{t}_0)) \cdot \int_{\mathrm{t}_0}^{\mathrm{t}} \ln [i(\mathrm{t}) - (1/p)(dp/d\mathrm{t})] d\mathrm{t} \right\} \tag{4'}$$

This formula is shown but not derived in the classic, *Inequalities*, by Hardy, Littlewood, and Polya, p. 136. The derivation requires only elementary calculus.

[2] The reader who checks my price level data against the source materials will find my numbers different from those published. None of the sources gives a single *GNP* deflator series for 1900–68. Series E-13 is used for 1899–1929 to get price changes for 1900–29. I linked E-13 and F-5 at 1929. I first linked the series using 1958=100. The level shown in the sources are averages for the entire year. I have thought of these numbers as June-July price levels. The Durand series uses prices from the first quarter, on average I think of these prices as mid-February ones. Then in order to have the price level for a given year t reflect price changes in the twelve months preceding the date of the bond quotation, I calculated and show as price levels $P'_{\mathrm{t}} = .375P_{\mathrm{t}-1} + .625P_{\mathrm{t}}$ where P_{t} are the published figures. I have also performed all calculations reported here for the P_{t} as well as P'_{t} and will send the results to those who request them.

[3] The number for year t in column (2) is $\{P'_{\mathrm{t}}/29.0 \exp[g(\mathrm{t}-1900)]\}$ where $g = (1/68)\ln(P'_{1968}/P'_{1900}) = .0209$.

Table 1—Realized Interest Rates and Holding Period Yields

Year	Price Level and Changes			Realized Interest Rates			One Year Holding Period Returns		
	(1) Level of Price Index	(2) Price Index Movement Around Trend Line	(3) One-Year Percent Change in Price Index	(4) Durand One-year Yield to Maturity	(5) Anticipated Real One-year Rate $\frac{(4)-(3)}{1+.01(3)}$	(6) Actual Realized One-year Rate $\frac{(4)-(3_{t+1})}{1+.01(3_{t+1})}$	(7) S and P AAA Composite Yield to Maturity	(8) Nominal One-Year Holding Period Return on (7)	(9) Realized One-Year Holding Period Return $\frac{(8)-(3)}{1+.01(3)}$
1900	29.01	100.00	7.53	3.97	−3.31	2.17	4.44		
1901	29.52	99.65	1.76	3.25	1.46	−0.25	4.39	5.15	3.33
1902	30.56	101.01	3.51	3.30	−0.20	0.18	4.39	4.44	0.90
1903	31.51	101.99	3.11	3.45	0.33	2.88	4.61	1.59	−1.47
1904	31.68	100.42	0.55	3.60	3.03	3.11	4.36	8.01	7.42
1905	31.83	98.81	0.48	3.50	3.01	1.45	4.27	5.61	5.11
1906	32.47	98.72	2.02	4.75	2.68	0.22	4.45	1.97	−0.05
1907	33.94	101.04	4.52	4.87	0.33	5.13	5.05	−3.02	−7.21
1908	33.86	98.70	−0.25	5.10	5.36	1.79	4.53	12.13	12.41
1909	34.96	99.80	3.25	4.03	0.76	−1.24	4.52	4.71	1.41
1910	36.82	102.94	5.34	4.25	−1.03	7.96	4.60	3.53	−1.72
1911	35.55	97.34	−3.44	4.09	7.80	3.22	4.59	4.78	8.51
1912	35.85	96.12	0.84	4.04	3.17	1.03	4.65	3.86	2.99
1913	36.92	96.93	2.98	4.74	1.71	5.96	4.89	1.63	−1.31
1914	36.50	93.83	−1.15	4.64	5.86	4.28	4.90	4.82	6.04
1915	36.62	92.20	0.35	4.47	4.11	−9.36	4.73	7.20	6.83
1916	42.21	104.07	15.26	3.48	−10.22	−22.02	4.69	5.31	−8.63
1917	56.02	135.24	32.70	4.05	−21.59	−12.97	5.42	−4.24	−27.84
1918	66.97	158.33	19.55	5.48	−11.77	−2.08	5.32	6.78	−10.68
1919	72.14	167.02	7.72	5.58	−1.99	−3.39	5.82	−0.76	−7.87
1920	78.83	178.74	9.28	6.11	−2.90	32.50	6.33	−0.26	−8.73
1921	63.13	140.17	−19.92	6.94	33.54	31.04	5.45	16.85	45.92
1922	51.52	112.03	−18.39	5.31	29.04	3.08	5.10	11.42	36.53
1923	52.64	112.08	2.16	5.01	2.79	5.12	5.23	3.51	1.32
1924	52.58	109.64	−0.10	5.02	5.13	2.50	5.00	8.31	8.42
1925	53.88	110.01	2.46	3.85	1.36	4.02	4.88	6.63	4.07
1926	53.79	107.56	−0.16	4.40	4.57	8.90	4.71	7.18	7.35
1927	51.57	100.98	−4.13	4.30	8.79	5.29	4.56	6.75	11.35
1928	51.08	97.96	−0.94	4.05	5.04	4.47	4.74	2.28	3.25
1929	50.88	95.54	−0.40	5.27	5.69	7.57	4.99	1.60	2.01
1930	49.79	91.55	−2.14	4.40	6.68	11.81	4.59	10.41	12.82
1931	46.49	83.71	−6.63	3.05	10.37	14.26	4.95	0.06	7.17
1932	41.92	73.93	−9.81	3.99	15.30	10.00	4.97	4.75	16.14
1933	39.64	68.45	−5.46	2.60	8.53	−1.08	4.70	8.61	14.88
1934	41.11	69.52	3.72	2.62	−1.06	−0.61	3.94	15.34	11.20
1935	42.45	70.30	3.25	1.05	−2.13	0.55	3.51	9.94	6.48
1936	42.66	69.19	0.50	0.61	0.11	−2.05	3.21	7.71	7.71
1937	43.82	69.60	2.72	0.69	−1.98	0.01	3.27	2.42	−0.29
1938	44.12	68.62	0.68	0.85	0.17	2.39	3.09	5.79	5.08
1939	43.46	66.19	−1.50	0.57	2.10	0.17	2.97	4.78	6.38
1940	43.64	65.08	0.40	0.41	0.01	−4.67	2.84	4.80	4.38
1941	45.96	67.13	5.33	0.41	−4.67	−9.20	2.86	2.60	−2.59
1942	50.82	72.69	10.58	0.81	−8.84	−7.47	2.86	2.89	−6.95
1943	55.37	77.55	8.95	1.17	−7.14	−2.86	2.83	3.29	−5.20
1944	57.67	79.10	4.15	1.08	−2.95	−1.42	2.73	4.18	0.03
1945	59.14	79.43	2.54	1.02	−1.48	−6.77	2.58	4.84	2.24
1946	64.07	84.27	8.35	0.86	−6.91	−9.79	2.58	2.63	−5.28

TABLE 1—REALIZED INTEREST RATES AND HOLDING PERIOD YIELDS—*Continued*

Year	Price Level and Changes			Realized Interest Rates			One Year Holding Period Returns		
	(1) Level of Price Index	(2) Price Index Movement Around Trend Line	(3) One-Year Percent Change in Price Index	(4) Durand One-Year Yield to Maturity	(5) Anticipated Real One-year Rate $\frac{(4)-(3)}{1+.01(3)}$	(6) Actual Realized One-year Rate $\frac{(4)-(3_{t+1})}{1+.01(3_{t+1})}$	(7) S and P AAA Composite Yield to Maturity	(8) Nominal One-Year Holding Period Return on (7)	(9) Realized One-year Holding Period Return $\frac{(8)-(3)}{1+.01(3)}$
1947	71.64	92.27	11.80	1.05	−9.62	−6.87	2.84	−0.95	−11.40
1948	77.72	98.03	8.50	1.60	−6.36	−0.40	2.77	3.90	−4.24
1949	79.29	97.93	2.01	1.08	−0.91	0.45	2.55	5.87	3.78
1950	79.79	96.51	0.63	1.42	0.79	−3.18	2.63	1.50	0.86
1951	83.57	98.99	4.75	2.05	−2.58	−1.72	3.05	−3.10	−7.49
1952	86.79	100.67	3.84	2.73	−1.07	1.31	2.99	3.92	0.08
1953	88.00	99.96	1.40	2.62	1.20	1.34	3.08	1.84	0.43
1954	89.11	99.12	1.26	2.40	1.13	0.93	2.89	5.72	4.40
1955	90.41	98.48	1.46	2.60	1.12	−0.08	3.16	−0.68	−2.11
1956	92.84	99.03	2.68	2.70	0.02	−0.88	3.84	−5.69	−8.15
1957	96.19	100.48	3.61	3.50	−0.11	0.50	3.66	6.36	2.65
1958	99.06	101.33	2.99	3.21	0.21	1.23	4.11	−2.17	−5.01
1959	101.00	101.17	1.96	3.67	1.68	1.99	4.61	−2.20	−4.08
1960	102.66	100.71	1.65	4.95	3.25	3.49	4.37	7.93	6.18
1961	104.11	100.01	1.41	3.10	1.67	1.89	4.43	3.55	2.11
1962	105.35	99.10	1.19	3.50	2.28	2.21	4.19	7.71	6.44
1963	106.67	98.27	1.26	3.25	1.97	1.73	4.36	2.09	0.82
1964	108.26	97.67	1.49	4.00	2.47	2.22	4.35	4.47	2.94
1965	110.15	97.31	1.74	4.15	2.37	1.73	4.72	−0.31	−2.01
1966	112.77	97.56	2.38	5.00	2.56	1.89	5.39	−3.47	−5.71
1967	116.21	98.45	3.05	5.29	2.17	1.51	6.12	−3.37	−6.23
1968	120.54	100.00	3.72	6.24	2.43		6.51	1.54	−2.10

Source: (1) Series E-13 and F-5 of U.S. Bureau of Census, *Historical Statistics*. Changed as explained in footnote 2 of text. (2) See footnote 3 of text. (3) Calculated. (4) *Economic Almanac*. (5)–(6) Calculated. (7) Standard and Poor's. (8)–(9) Calculated.

The numbers in column (2) are not used in any calculation of realized rates.

The percent price change for year t, $(p_t - p_{t-1})/p_{t-1}$, is shown in column (3). For example, the entry for 1960 shows the percent rise in the price level from mid-February 1959 until mid-February 1960.[4]

[4] Fn. 2 contains the rationale for and the details of constructing the price levels on a mid-February basis. There is no compelling reason to use 12 rather than some other number of months in the calculation of the rate of increase of prices for the anticipated realized rate series. The St. Louis Federal Reserve Bank has used a twenty-four month approximation for a similar purpose; see their *Monthly Trends*. I have chosen to use a 12-month span because that is the same length of span I use to calculate the actual realized rates.

To calculate the anticipated real rate, the percent price change over the previous 12 months is subtracted from the quotation of the nominal rate on a given date and the difference is divided by a correction factor. The number shown in column (5) of Table 1 for a given year shows the anticipated real rate for the year following the rate quotation. Thus the number for year t in column (5) is obtained by subtracting the number in column (3) for year t from the number shown in column (4) for year t and dividing the difference by (1+number in column (3)). The assumption that last year's price change will persist is not, of course, the only assumption that may be made about price expectations.

Indeed, the assumption of a one-year persistence may even be unreasonable since it results in negative anticipated realized rates for 23 of the 68 years shown. For *any* constant n and these data, assuming n-year price change persistence will lead to some negative anticipated prices—clear evidence that a theory of anticipated real interest rates must incorporate estimates of price changes over time spans of varying length, or that these data or my use of them is faulty.

The actual or realized rate is shown in column (6) of Table 1. The entry for year t is the number in column (4) for year t less the number in column (3) for year t+1 divided by (1+number in column (3), year t=1). Thus the number in column (6) shows a rate quotation for a year less the price change in the 12 months that followed that quotation divided by a correction term.

The number shown in column (5) is a forecast based on the assumption that the past rate of price increase will persist. This number is known at the same time that the nominal rate quotation in column (3) is known. The number shown in column (6) better measures what happened, but it cannot be calculated until after the fact.

One may not conclude from these data that the nominal rate is empirically well approximated as the sum of a "real" rate constant over time and the relative change in prices. If it were, the plot or regression of the nominal rate, column (4), against the percent change in prices, column (3), would be significant. It is not—the plot is essentially circular; in the regression, the nominal rate is a function only of the constant term and even then the fit is virtually nonexistent. This regression is presented, with others, in footnote 8 below. The regression results were not significantly improved by using as the independent variable the percent price change for any 12-month period adjacent to or overlapping the date of nominal rate quotation.

So that the reader may more easily think about procedures different from mine, I have provided column (2) in Table 1 and Figure 1. The numbers in column (2) help one ascertain abnormal price movements. In Figure 1, I have plotted *GNP* price-deflator movements with trend removed, along with the nominal interest rates from columns (4) and (7). It is clear from Figure 1 that "high" prices, relative to the trend line, correspond with high interest rates and similarly for low prices and rates.

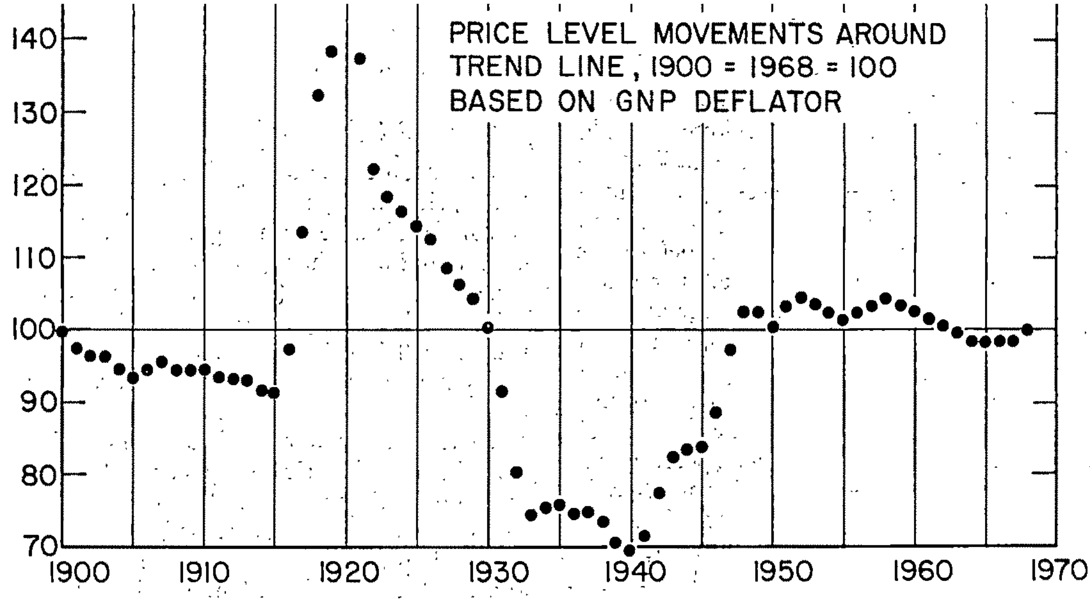
II. *Holding Period Returns to Bonds*

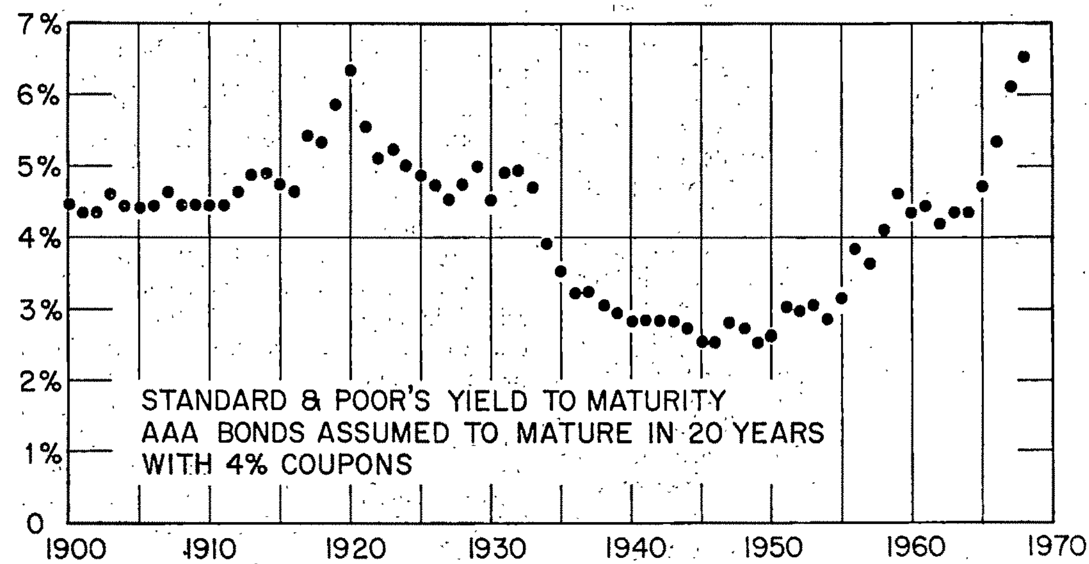
While yield to maturity data corrected for price changes may be confidently used to measure realized interest rates, they cannot measure returns to bondholders unless time to maturity is the same as the holding period for which a return is required. This difference occurs because part of the typical yield is a capital gain paid when the bond matures. Suppose, for example, a bond which currently sells for \$90 matures in exactly 2 years with principal repayment of \$100 and in addition pays interest of \$5 at the end of each of the next two years. The *yield to maturity* as shown in a typical series would be 10.8 percent because

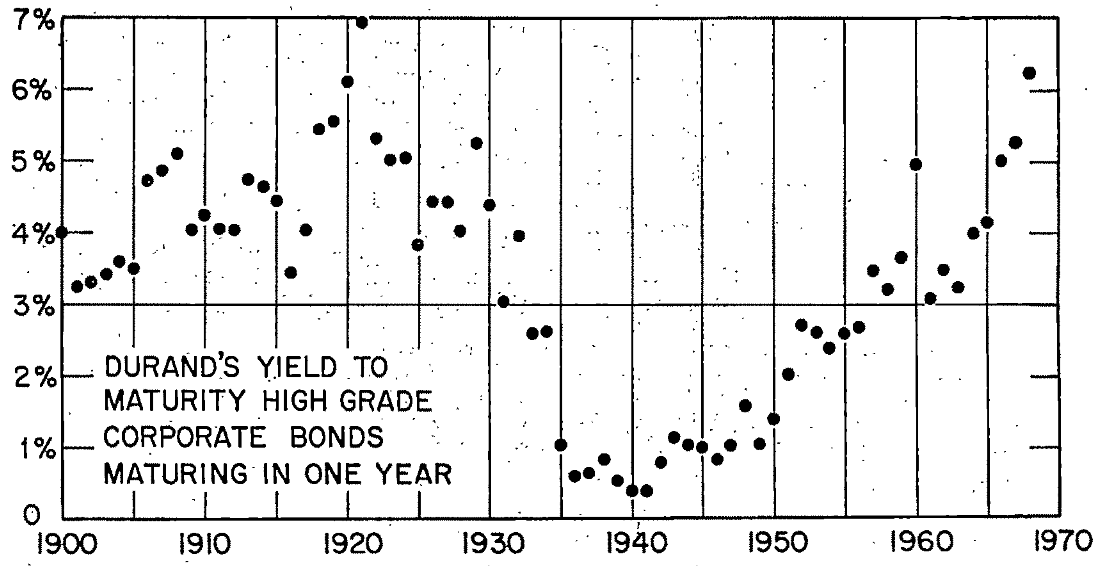
$$(5) \qquad 90 = \frac{5}{1.108} + \frac{105}{(1.108)^2}$$

The *realized rate of return* at the end of one year would be 10.8 percent if, and only if, the price of the bond one year hence were \$94.74 plus accrued interest of \$5 (.108=(94.74+5−90)/90).

As the actual value one year hence varies from below to above \$99.74, the realized return would vary from below to above 10.8 percent but the yield to maturity quoted one year hence would vary inversely, from above to below 10.8 percent.


PRICE LEVEL MOVEMENTS AROUND
TREND LINE, 1900 = 1968 = 100
BASED ON GNP DEFLATOR
140
130
120
110
100
90
80
70
1900
1910
1920
1930
1940
1950
1960
1970


STANDARD & POOR'S YIELD TO MATURITY
AAA BONDS ASSUMED TO MATURE IN 20 YEARS
WITH 4% COUPONS
7%
6%
5%
4%
3%
2%
1%
0
1900
1910
1920
1930
1940
1950
1960
1970


DURAND'S YIELD TO
MATURITY HIGH GRADE
CORPORATE BONDS
MATURING IN ONE YEAR
7%
6%
5%
4%
3%
2%
1%
0
1900
1910
1920
1930
1940
1950
1960
1970

The rate of return from holding a bond with the yearly coupon payments for one year is $[P(t)-P(t-1)]/P(t-1)$ assuming the price $P(t-1)$ is measured just after an interest payment and $P(t)$ just before the next one. Since the term structure is not flat and future interest rates are uncertain, the holding period return for a given time span measured from bonds to mature at the end of that span will be different from the return from bonds that do not mature at the end of the period. The only way an investor can assure his return (even assuming no default risk) is for him both to know his horizon and to buy a bond that matures at that horizon. Otherwise he is subject to capital gains (or losses) caused by decreases (or increases) in the rate of interest.[5] One may make various assumptions concerning the horizon of the buyer and the maturity he buys. For the purposes of this study I have assumed a holder of long-term, high-grade corporate bonds who alters his portfolio each year (with no transactions costs) to maintain a term to maturity of approximately 20 years. It is perhaps true that no *individual* bondholder behaves much this way but I believe it reasonable to suppose that an important *class* of bondholders do. Thus the series I develop is of interest to those curious about the yearly returns of the long-term bond-holder whose portfolio on average contains 4 percent, 20-year bonds. Alternatively, my series shows the average returns in the first year of ownership to buying a high-grade, 20-year maturity, corporate bond portfolio.

[5] As a matter of fact, buying bonds which mature at the horizon will not assure the return. The *promised* return at time of initial investment is the yield to maturity and assumes reinvestment of coupon payments at that same yield. The true return will depend upon the yields to maturity at times of reinvestment of the coupon payments. Thus a policy of buying and reinvesting in the bond of horizon maturity cannot assure, in the strictest sense of the word, the return. The strategy of buying bonds of horizon maturity is inferior to one of buying bonds, if necessary altering the portfolio, so that the duration (Macaulay, pp. 44–53) of the portfolio is maintained equal to the amount of time between the reinvestment date and the horizon. Lawrence Fisher and I, in a study now available, have implemented the duration strategy and for the first time have measured its superiority over maturity strategies on corporate bond data from 1925 through 1968.

Data and Results.

To measure holding period yields actual bond prices and coupon payments are analyzed. Reasonable data are available. Standard and Poor's has compiled a composite series of high-grade (AAA) corporate bond prices and yields for 1900–68. These are shown in Table 1, column (7). From 1900 through mid-1937, the price data are averages of actual bond prices from which are calculated the average yield to maturity; from 1937 on the prices shown are calculated from observed average yields to maturity.[6] For both sets of calculations the bonds are assumed to have carried 4 percent coupons and 20-year maturities. The bonds included in the portfolio are changed over time, thereby introducing a bias to be discussed below. I have used the data on average yield to maturity and the 4 percent, 20-year assumptions to calculate a series of one-year holding period nominal and realized returns. These results are shown in Table 1, columns (7), (8), and (9).

I have assumed, as does Standard and Poor's, that the coupon payments are received in semiannual \$2 installments and that the payment is reinvested in the bond series at the current price. Thus if the prices at the end of two consecutive years are P_1 and P_2, and if the value of the coupon payments made during the year and accumulated at compound interest is V, then the holding period yield H is defined by

$$H = (P_2 + V - P_1)/P_1 \tag{6}$$

If bond coupon payments were made only once a year, V would equal the coupon payment or, in this study, \$4 per \$100 bond. But the coupon payments are made twice a year so that

$$V = 2(1 + r) + 2 \tag{7}$$

where r is the rate at which the \$2 coupon payment is reinvested. I assume the price of the 20-year AAA bond changed uniformly

[6] From 1900 through 1940, I have used the S & P average of one price from each week in December. From 1941 to 1968, I have used the price from the last week of December, a better procedure for this kind of calculation.

from P_1 to P_2 and that the coupon payment was reinvested in that bond at midyear so

$$(8) \qquad 1 + r = (P_2/P_1)^{1/2}$$

In calculating the entry for year t in column (8) of Table 1, I used equation (6) where P_1 is the price in year t−1 of a 4 percent bond maturing in 20 years with current yield to maturity as shown in row t−1 of column (6). P_2 in equation (6) is the price of a 4 percent bond maturing 19 years hence with yield to maturity from row t of column (6).

The sample of bonds used by Standard and Poor's biases upwards the calculated returns. The sample group changes over time. As a bond became unworthy of a AAA rating, it would be removed from the sample. It seems reasonable to suppose that some of the bonds whose quality has worsened continue to worsen until they default. Thus the composite sample contains fewer bonds which eventually default than would an individual's portfolio of AAA bonds started in 1900 and appropriately added to until 1968. I argue, then, that the estimates shown in Table 1, column (8), if inaccurate because composite bond data are used, overstate actual *one-year holding* period yields on long-term bonds.

Another bias, the effect of which is harder to trace, is introduced because the composite bond sample does not always consist of bonds with $4 coupon payments and 20 years until maturity. The yields to maturity shown in column (7) of Table 1 could have been calculated from bonds with, say, $5 coupon payments and 15 years to maturity. Then the average one-year holding period return measured over 1900–68 would have been 3.94 percent rather than 3.91 percent as calculated assuming 4 percent, 20-year bonds. In order to establish the order of magnitude of this possible bias, I have constructed bond price series and corresponding one-year holding period returns from the yield-to-maturity data as shown in column (7) of Table 1 but assuming, in turn, $3, $4, and $5 coupon payments and 15, 20, and 25 years until maturity. The mean holding period yields and their standard errors from these nine pairs of assumptions are shown in Table 2. The measured returns and their means do not appear significantly different for varying assumptions about coupons and maturities. The pattern of mean returns in each row are consistent with predictions derived from analysis of Macaulay's (pp. 48f) concept of duration.[7] For a given coupon, the longer the term to maturity, the longer is the duration of the bond. The longer the duration of a bond, the less valuable is the bond when nominal interest rates rise as they have from 1900–68. The numbers in each column are essentially constant, but insofar as they differ the pattern is predicted by analysis of the duration which is shorter, the larger the coupon for a given maturity.

Table 2—Mean Nominal One-Year Holding Period Yields, 1901–1968 (and Standard Error) From Yield-to-Maturity Data Shown in Table 1 for Varying Coupons and Maturities

(Mean Realized Returns Are in Each Case 2.34 Smaller)

Percent	15 Years	20 Years	25 Years
3	3.96 (.50)	3.92 (.56)	3.89 (.64)
4	3.97 (.44)	3.93 (.53)	3.91 (.60)
5	3.98 (.42)	3.94 (.51)	3.92 (.58)

The years-to-maturity assumption appears more critical than the coupon assumption. The relative importance of the two assumptions depends, in practice, on what is happening to underlying real rates and to the term structure. I am unable to say whether assumptions made about the coupon and term to maturity bias the estimates of holding period returns. It seems fair to conclude, since the measured mean rate of price increase was 2.42 percent, that average annual realized bondholders' returns on high grade corporate bonds have been less than 1.5 percent in the twentieth century. The mean realized holding period returns in 1901–10 were 2.51 percent with standard error 1.65

[7] If P_t is the present value of the payment promised for the end of period t, then that stream of promised payments has *duration* equal to $\Sigma tP_t/\Sigma P_t$.

percent and in 1959–68 were −0.20 percent with standard error 1.48 percent. Note that the time span (mid-February of one year to the next) over which each price increase is calculated and the time span (late December of one year to the next) over which each holding period return is calculated are not exactly the same.

III. *Conclusion*

Yield to maturity on a bond corrected for price change is a reasonable estimate of the actual or realized interest rate. If in the long run anticipations are correct the estimate of realized rates approximate anticipated real rates as well.

Further I infer from the data that on average for 1900–68 expectations or forecasts of annual price change were substantially incorrect or that there was no constant "real" rate of interest, or both. Had expectations been correct and had there been a real constant rate of interest, then there should be a significant linear relationship of the form

$$\text{(nominal rate)} = a + b(\text{relative price change}) \tag{9}$$

where a would be that real constant rate of interest and b would be about one. There was no such significant relationship over the entire period. My opinion is that short-run expectations were faulty. This opinion is reinforced by the difference between one-year interest rates, Table 1, column (4), and one-year bondholders' returns, Table 1, column (8). One explanation for this difference is that anticipations are biased. The relation in equation (9) between the nominal rate and the relative price change does appear significant for recent years so that the St. Louis Federal Reserve Bank has been able to plot an essentially constant real rate of interest equal to the market rate on corporate bonds less the rate of change in prices over the preceding two years.

The regression results for equation (9) fitted for two definitions of the nominal rate and two time spans are as follows. Standard errors are shown in parentheses.

Durand	R^2	a	b
1. 1900–68	.030	3.449(.208)	−0.041(.029)
2. 1961–68	.936	1.990(.271)	1.146(.123)
Standard and Poor's			
1. 1900–68	.007	4.245(.128)	−0.013(.017)
2. 1961–68	.989	3.075(.091)	0.953(.041)

The nominal rate is Durand's one-year rate, column (4) of Table 1, in the first two regressions and Standard and Poor's yield to maturity, column (7) of Table 1, for the second two. The relative price changes are from column (3) of Table 1.

The yield to maturity of a bond gives little information about rate of return to holding bonds unless the holding period is the same as the term to maturity.[8] The one-year return to investors who maintain a portfolio of high-grade corporate bonds with yield to maturity of roughly 20 years was 3.93 percent annually (1.51 percent after corrections for price changes). In a period of rising interest rates investors maintaining a portfolio of constant maturity or duration would have been better off choosing any term shorter than 20 years, the shorter the better.[9]

References

D. Durand, *Basic Yields of Corporate Bonds.* New York 1942.

I. Fisher, *The Theory of Interest.* New York 1930.

L. Fisher and R. L. Weil, "Policies for Minimizing Uncertainty of Bond Investments," Center for Mathematical Studies in Business and Economics, Report 7030, Univ. Chicago.

W. E. Gibson, "Price-Expectation Effects on Interest Rates," *J. Finance,* Mar. 1970, *25,* 19–34.

G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities.* Cambridge 1934.

F. R. Macaulay, *Some Theoretical Problems*

[8] The correlation between the numbers in columns (4) and (8) of Table 1 is .06, not statistically different from zero.

[9] After this article was written, two articles appeared by Gibson and by Yohe and Karnosky. None of those results have been incorporated here.

Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856. New York 1938.

W. P. Yohe and D. S. Karnosky, "Interest Rates and Price Level Changes, 1952–69," *Fed. Reserve Bank St. Louis Rev.*, Dec. 1969, *52*, 18–36.

National Industrial Conference Board, P. Biederman, ed., *Economic Almanac, 1967–1968 Business Fact Book.* New York 1967.

St. Louis Federal Reserve Bank, *Monetary Trends* (monthly). See, for example, January 1969.

Standard and Poor's. *Trade and Securities Statistics: Security Price Index Record.* New York 1968.

U.S. Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1957; Continuation and Revisions.* Washington 1961, 1965.

The Efficient Production of External Economies

By MANCUR OLSON, JR. AND RICHARD ZECKHAUSER*

Does independent market adjustment always lead to a less than Pareto optimal supply of goods when there are external economies? The traditional Pigovian conclusion that it did has recently been rejected by some eminent economists. This rejection began with a contribution in this *Review* by James Buchanan and Milton Kafoglis, who produced examples in which independent market adjustment apparently provided optimal or even supra-optimal supplies of goods with external economies, and thus found the orthodox Pigovian conclusion unsatisfactory. Their finding has been accepted by many economists, including William Baumol, who attempted, in a subsequent communication in this journal, to provide a formal, general explanation of their result. This note will show that Buchanan and Kafoglis failed to call attention to the major theoretical point suggested by their examples, and accordingly misinterpreted them. This theoretical point, which is our principal concern, has important implications for economic policy.[1]

The finding that independent market organization of an activity with external economies need not lead to an undersupply of resources for that activity seemed, to Buchanan and Kafoglis, Baumol, and others, to raise grave new doubts about the case for collective intervention in markets with externalities.[2] In fact, it has exactly the opposite implication. This is evident from an understanding of the theoretical point, overlooked in the literature, that different individuals generate external economies with different degrees of efficiency. Independent adjustment in the presence of externalities will generally lead not only to an inefficient level of production, but to an inefficient method of production as well.

A proper understanding of one of Buchanan and Kafoglis' simplest examples will make all of this clear. They consider a hypothetical situation in which immunization for one individual, B, provides an external economy to another individual A, but the relationship is "nonreciprocal," in that immunization for A does not benefit B. Since income effects are ignored, Buchanan and Kafoglis can use marginal evaluation curves as demand curves, and describe the situation with the simple geometric presentation of Figure 1. The valuation individual A places upon protection from contagious disease at the margin is given by demand curve D_A in Figure 1A, and the same is true of B's demand curve D_B in Figure 1B. The marginal cost of shots, represented by the CC lines, is assumed constant and equal for both parties. Individual A, in isolation, would purchase OQ of disease protection, and B, who never

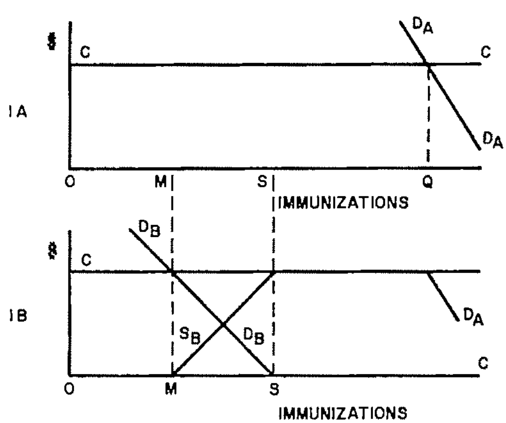


FIGURE 1

* William Baumol, James Buchanan, and Wallace Oates provided most helpful criticisms, but share no responsibility for the faults or conclusions of this note. The authors are, respectively, associate professor at the University of Maryland and assistant professor at Harvard University.

[1] In their 1967 article, the authors discuss the importance of this point and its implications in the context of international alliances for defense.

[2] Buchanan and Kafoglis were not entirely explicit about what they thought the policy implications of their argument were, and might have been misunderstood. See Buchanan and Gordon Tullock, Baumol, and fn. 4 below.

receives any spillover, would purchase OM, whether in isolation or not. When the nonreciprocal externality is taken into account, A's purchase will decline. If B's shots are a perfect substitute for his own, A will reduce his purchase by one unit for every unit that B has obtained. In other words, he will reduce his purchase by an amount equal to OM, and purchase only MQ himself. The combined outlay, $OM+MQ$, is the total equilibrium outlay obtained through independent adjustment.

Buchanan and Kafoglis compare this market equilibrium with the Pareto optimal level of provision. Since each of B's shots does double duty, protecting A as well as himself, both individuals can be made better off by taxing A and using the proceeds to induce B to buy additional shots. The value to A of immunizations of B is given by the marginal cost curve CC until it cuts A's demand curve, and from this point on that demand curve indicates the value to A of additional immunizations for B. By subtracting B's demand curve from the marginal cost curve, we get S_B, B's supply curve of immunization protection for A. The ideal collective outlay is achieved where A's demand curve and B's supply curve intersect; that is, when B obtains OS of immunization and A purchases an additional SQ, for a total of OQ.

In this example, A's intake of shots is reduced by one for every extra shot B can be induced to take. The result is that the total number of shots taken is the same under an ideal (Pareto optimal) collective arrangement as it is under independent market adjustment. If an immunization shot for B were a more-than-perfect substitute for a shot for A, each extra shot B would take under an ideal collective arrangement would make it appropriate to reduce A's outlay by more than one shot, and the collective outlay on immunization would then be *less* than under independent adjustment. Buchanan and Kafoglis therefore draw the general conclusion that independent adjustment can lead to the same outlay, or even to a larger outlay, on a good with external economies than will ideal collectivization.

This example and its implications have been misunderstood. To gain a proper understanding, it is essential to consider the efficiency of the *method* of production of the external economy, and to distinguish it from the familiar question of whether there is an optimal or efficient *level* of production. The externality in the example was nonreciprocal. B's shots give immunity to A as well as himself, but A's shots protect him only. This means—and this is the simple essence of the example—that there are two different production functions or processes that can be used to produce immunity from disease in the example, and that these two productive processes differ in efficiency. If, as in the Buchanan-Kafoglis example, B's shots are a perfect or more-than-perfect substitute for shots for A, and B places a positive value on shots for himself, B is undoubtedly the more efficient producer of immunity. This can be seen in Figure 1, where it is evident that the output of immunity under the ideal collective arrangement $(OQ+OS)$ is larger than the output under independent market adjustment $(OQ+OM)$, even though the total number of shots taken is equal in the two cases. The reallocation of shots from A to B has increased total immunity.

In general, if there are two or more ways of producing a product and one method is more efficient than the others, it can easily be the case that more will be spent on the good if an inefficient method is used. (If the elasticity of demand for a good is less than unity over the relevant range, increasing the per unit cost of the good to its consumers will increase their expenditure on it, at the same time that it reduces the amount they consume.) Such differences in the efficiency with which external economies are produced explain all of the Buchanan-Kafoglis examples. Whenever one individual is a more efficient producer of an externality or collective good than another, ideal collectivization leads to a greater allocation of inputs to the more efficient producer, and more output per unit of input. The result is that expenditure on inputs may decrease with collectivization, but output increases.

As soon as it is understood that differences in the efficiency with which external economies are generated *entirely* explain the

Buchanan-Kafoglis findings, it becomes clear that Baumol's seemingly rigorous explanation and validation of their result is (though valuable for other purposes) irrelevant to this issue.[3] It also becomes clear that the policy implications have been misinterpreted. The Buchanan-Kafoglis paper leaves the impression that it calls into question "... the theory of economic policy upon which arguments for the collectivization of any activity must be based ..." and the "... orthodox policy implication ..." of that theory (p. 403). In fact, the major policy implication suggested by their examples, which they do not point out anywhere in their paper, is that the shortcomings of independent adjustment in a free market with externalities are greater than has previously been understood. There is now an additional reason for collective intervention when goods have external economies. Independent adjustment through *the market mechanism may not only provide less of the good than is optimal (as Pigou told us long ago), but will often also make society pay more for the suboptimal supply of the good than it would have had to pay for the larger amount an ideal collective arrangement would provide (as we have now discovered)*. In other words, it is now clear that independent adjustment in the presence of externalities will lead not only to the production of the wrong amount of a good, but also to an inefficient location or method of production. (There is an exception in the highly special case in which all parties who create external economies provide the spillover at the same cost and are thus equally efficient.) The Pigovian model overlooks the general connection between externalities and the method of production, and is to that extent inadequate, but the traditional conclusion that externalities lead to inefficient market outcomes is now twice blessed.[4]

The practical relevance of this point is evident from a glance at the real-world cases that Buchanan and Kafoglis discuss. They tell us that their interest in this problem arose out of an attempt to explain the fact that, contrary to what the Pigovian model would suggest, expenditures on medical services have expanded more rapidly in the United States than in Britain in the period since the British decided to collectivize medical care. The explanation of this phenomenon that is implicit in their examples is that the production of medical care is, because of nationalization, more efficient in Britain than in the United States, and that the British are getting more health protection than a comparison of expenditures on medical services would suggest. We may personally doubt that this is anything like a satisfactory explanation of an extremely complex phenomenon, but it is certainly the explanation implicit in the Buchanan and Kafoglis argument.[5] Their conjecture that total expenditure on guards, locks, firearms, and other forms of protection might increase if public provision of police services were to cease can be understood in the same way. Private, decentralized provision of police services would presumably be less efficient than public provision, with the result that the society might pay more for police protection and enjoy less.

At first glance, our analysis does not seem to be consistent with Buchanan and Kefoglis' example of reciprocal externalities. In this example, they use a payoff matrix to show the returns to individuals A and B for different numbers of immunization shots for each of them, taking account of the external economies that each provides to the other. In this way they show that *output* as well as *expenditure* on a good with external economies can be greater under independent adjustment than under an ideal collective arrangement. They argued that this result "... suggests that, under independent ad-

[3] The reasons why Baumol's argument is off the point, and the true implications of his analysis, are set out in our 1967 article.

[4] Nonetheless, in the real world collectivization may work so badly that the market distorted by externalities would often be the lesser evil.

[5] Elsewhere Buchanan deals with the National Health Service in more detail, and explicitly points to the logical possibility that increases in efficiency could account for the modest size of the increase in British medical expenditures since the National Health Service was established.

justment, the consumption of final output is overextended . . ." (p. 411). How can the logical possibility they have demonstrated by example be explained?

Once again the explanation centers on the differences in efficiency of different ways of generating external economies. It will here be necessary to distinguish between efficiency in the production of a given externality in a total or overall sense, and efficiency in the production of a particular marginal unit. Consider a case in which there are two relevant alternative ways to produce a given external economy or collective good. Method 1 is the manner that results from independent adjustment. Method 2 involves concentrating all of the generation of the collective good in the unit that can produce it at lowest total cost, as would be done under ideal collectivization. Though by assumption total costs of production (at any level of output) are lower using the second rather than the first method of production, it is nevertheless possible that, over some limited range, the *marginal* cost of producing the good is lower if method 1 is employed. This possibility is illustrated in Figure 2, where TC_1 shows the cost of producing the good via method 1 and TC_2 shows the cost under method 2. TC_2, the cost curve for an ideal collective arrangement, lies below TC_1. At point Q_2 marginal value and marginal cost (given by the slopes of the total valuation curve and TC_2) are equal, so OQ_2 is the optimal level of output. Had the less efficient method 1 been employed, however, marginal costs would have been lower over a considerable and relevant range. With TC_1, total gains are greatest at output OQ_1, or at an output greater than the Pareto optimal output. This shows that in certain special circumstances the most advantageous level of output will be larger when a less efficient method of production is used.[6]

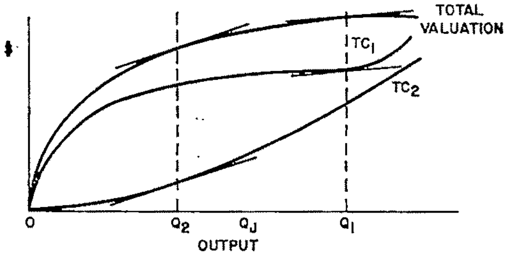


FIGURE 2

Independent adjustment with external economies would not, to be sure, lead to output OQ_1. For the standard Pigovian reasons, each party will ignore the benefits to others of the externality it generates and will produce too little. Independent adjustment will accordingly lead to an output such as OQ_J, which is less than that which maximizes the gain along cost function TC_1, but may well be greater than the Pareto optimal output OQ_2. This shows that it is possible, as in the Buchanan-Kafoglis example, that a larger *output* of an externality would result from independent adjustment than from a Pareto optimal collective arrangement.[7]

[6] Independent adjustment could *not* lead to a larger output than an ideal collectivity, even in these special circumstances, if the ideal collectivity could have produced *every* unit of the externality at a *marginal* cost as low or lower than that confronting any individual member. If the ideal collectivity could have used method 2 to produce all of the units that could have been produced with the least cost by that method, and also used method 1 when its marginal costs were lower (as they are over that portion of TC_1 which is flatter than the corresponding portion of TC_2), then the Buchanan-Kafoglis result in the text would not be possible. This is because the assumption that the collectivity enjoys marginal costs as low at every point as any member obviously insures that its costs of production could not rise faster than under a nonoptimal method of production. The special condition required for the Buchanan-Kafoglis result could, however, clearly be met in practice. This condition would be satisfied if the marginal cost of generating an external economy through a given individual were a function of the number of units *he* (rather than the collectivity) had previously produced.

[7] An examination of the Buchanan-Kafoglis payoff matrix makes clear that their result is in fact explained by our argument. Their Pareto optimal outcome consists in giving 3 shots to B and none to A, but under independent adjustment A and B each get two. When three shots have been given to B, an additional shot brings so little extra immunity it is not worth the cost, whereas when the generally less efficient practice of dividing the shots between the two is followed the fourth shot adds a great deal of immunity (i.e., has lower *marginal* costs in the relevant range).

The payoff matrix in the example also has two maxima. Social welfare functions which allow the result of

The fact that output under independent adjustment can be greater than under ideal collective arrangements does not, however, justify the Buchanan-Kafoglis claim "... that, under independent adjustment, the consumption of final output is overextended," even in the special case we have considered. If the inferior production processes used under independent adjustment continue to be used, every one can be made better off by increasing output still further. This is because individuals are not rewarded for the external economies they provide, and accordingly produce too little with any *given* method of production. The output under independent adjustment cannot then be said to be overextended since if it were reduced and no other changes were made, the outcome would be worse for all concerned.

The fundamental concern here is not with the errors and findings in earlier writings, but with the theoretical and policy insights that a proper understanding of this issue allows. The fact that different sources of external economies or collective goods normally generate them with different degrees of efficiency suggests that the production of and "trade" in spillovers should promote productive efficiency by following the principle of comparative advantage. The relatively most efficient producers of a given external economy should "specialize" in that externality, and receive different spillovers or ordinary goods in return. This means that there is some theoretical justification for giving the largest scholarships to the brightest students, for concentrating public health programs on those groups most susceptible to contagious disease, for giving television broadcasting licenses to those stations with the tallest transmitters. In a defense alliance, where each nation's military expenditure provides an external economy to its allies, each member nation could specialize in providing the type of military strength in which it had comparative advantage (e.g. in NATO, the United States could provide missiles, Britain ships, Turkey infantry, etc., as explained in Olson and Zeckhauser, 1967). Those allies whose comparative advantage was not in any type of military capability could reduce their military expenditure and send their allies civilian goods in return for military protection. Similarly, the large city center government in a metropolitan government could provide a large library and a police communications system for the entire metropolis, and the spacious suburbs the land for the airport and the city dump.

Though the literature on this issue has ignored external diseconomies, the argument here also applies to (and has policy implications for) them. When firms producing a good generate external diseconomies, independent adjustment will mean not only that too much of this good is produced, but also that the distribution of output among the firms in the industry, and the nature of efforts to minimize the damage done by the diseconomies, will generally be inefficient. As the analogy with external economies correctly suggests, the firms with the largest external diseconomies per unit of output should restrict output the most, since (other things equal) they can deal with the diseconomy with the least cost. But the emphasis on differences in efficiency in dealing with externalities here calls attention to another possibility; sometimes the most efficient way to limit the damage done by diseconomies is not to limit or change the activity of the offending parties, but rather to have the victims of the diseconomy deal with it. The cost of moving away from or otherwise adjusting to the diseconomies may be less than the costs of preventing or limiting the diseconomy at its source. This in turn means that the commonplace suggestion that those who generate external diseconomies ought to have to compensate their victims for any losses they suffer, can work against Pareto optimality. When such a suggestion is adopted, those injured by the diseconomy

greater output under independent adjustment will in general have two or more maxima, because there must be a separate local maximum accounting for the greater output under independent adjustment. Welfare functions with multiple maxima and areas where the second order conditions for a maximum are not fulfilled are, as Baumol correctly pointed out, more likely when externalities are present.

have no incentive to protect themselves from it, even if this should be more economical than requiring adjustments on the part of those who generate the diseconomy.

Thus, whether external economies or external diseconomies are at issue, independent market adjustment will in general be inefficient in two distinct ways, only one of which has been properly understood. When there are externalities of either kind, it will not only be true that independent adjustment in a free market will lead to production at a level that is not Pareto optimal, but also that the method of production of the external economy or the effort to combat the external diseconomy will in general be inefficient. In other words, the location of the needed efforts will not be in accord with the principles of comparative advantage. It follows that, contrary to the impression created by the recent literature, what we have is a new idea about how resources should be allocated in the presence of externalities, and an additional argument for collective intervention in markets with externalities. Whether, in view of the often striking shortcoming of collective institutions in the real world, the new argument for intervention should be decisive is another question. But it is surely important that it be understood.

REFERENCES

W. Baumol, "External Economies and Second Order Optimality Conditions," *Amer. Econ. Rev.*, June 1964, *54*, 358–72.

J. Buchanan, *The Inconsistencies of the National Health Service*, London 1965.

——— and M. Kafoglis, "A Note on Public Goods Supply," *Amer. Econ. Rev.*, June 1963, *53*, 403–14.

——— and G. Tullock, "Public and Private Interaction Under Reciprocal Externality," in J. Margolis, ed., *The Public Economy of Urban Communities*, Baltimore 1956, pp. 52–73.

M. Olson, Jr., *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, Mass. 1965.

——— and R. Zeckhauser, "An Economic Theory of Alliances," *Rev. Econ. Stat.*, Aug. 1966, *48*, 266–79

——— and ———, "Collective Goods, Comparative Advantage, and Alliance Efficiency," in R. McKean, ed., *Issues in Defense Economics*, Nat. Bur. Econ. Res., New York 1967.

A Simple Model of Replenishable Natural Resource Exploitation

By C. G. Plourde*

The purpose of this paper is to consider maximum sustained yield programs of replenishable natural resource exploitation and to question the validity of these programs in satisfying social goals.

In the treatment of such programs, two fundamental problems arise. The first is that the existence of a social discount factor (or interest rate) may cause the maximum sustained yield program to be nonoptimal.[1] The second problem, considered recently in the literature, relates to the many externalities which may be present in harvesting resources.[2]

The most significant of these externalities is the 'stock' externality in production. That is, there is a potential misallocation of inputs in the production of natural resource product due to the fact that one input, the natural resource itself, contributes to production but may not receive payment (for example, its marginal product) because no one owns the resource. Recently, research in this area has been directed at finding optimal taxing schemes which will have the effect of assigning to an unappropriated natural resource its imputed rent.[3]

In this paper, a simple model will be developed which will illustrate the nature of the first of these problems in a case where the second problem is not present. To accomplish this, production will be assumed 'costless' in the sense that no inputs are used. (Production, which will be the same as consumption, has a 'cost' in the sense that it diminishes the stock of resource available for consumption in all subsequent periods.)

The use of a natural resource in this model is parallel in theory to the use of capital, and the term 'investment' will be used here to indicate simply 'foregone consumption.'

The intertemporal aspect of the problem suggests a dynamic model, and the model is formulated as an optimal control question. The approach parallels Kenneth Arrow in his treatment of the Reversible Ramsey Problem.

Population Dynamics of a Replenishable Natural Resource

Following A. Lotka, V. Volterra, and others, the equation

$$\frac{dN_t}{dt} = \lambda N_t - \epsilon N_t^2 \tag{1}$$

will be assumed to represent the population growth of a natural resource over time.[4] λ and ϵ are given parameters, assumed constant over time, and N_t represents the mass of natural resource at time t. The graph of dN_t/dt is given in Figure 1. The solution of (1) is a logistic curve.

When man is introduced (as a predator) it is assumed that he consumes C_t units of natural resource at time t, and that the growth law becomes:

* The author is assistant professor at the University of Western Ontario. This paper is essentially one chapter of a Ph.D. dissertation written at the University of Minnesota and supported by the Canada Council. The author wishes to thank his adviser Professor Edward Foster and other members of the economics department at Minnesota for their assistance. Any errors are solely the responsibility of the author.

[1] H. S. Gordon and later J. Crutchfield and A. Zellner treated this problem in detail from the viewpoint of production efficiency.

[2] See A. Scott, V. Smith (1968, 1969), J. Quirk and V. Smith, and Crutchfield and Zellner.

[3] Smith (1968) considers optimal taxing schemes, as do Quirk and Smith. I have also considered this problem, as well as an optimal quota system, in my unpublished Ph.D. dissertation.

[4] Some authors, notably Smith, use a more general specification of dN/dt. The above specification is chosen to better illustrate the relationship between optimal sizes of N_t and maximum sustained yield sizes.

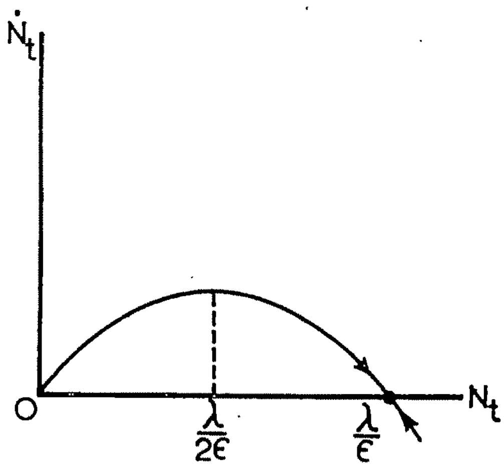


FIGURE 1

(2) $$\frac{dN_t}{dt} = \lambda N_t - \epsilon N_t^2 - C_t$$

A Simple General Equilibrium Model Involving a Replenishable Natural Resource

Suppose a planner wishes to control the consumption of natural resources so as to maximize the welfare functional

(3) $$\int_0^\infty U(C_t)\cdot e^{-\delta t}dt$$

where it is assumed that:

i) human population is constant;
ii) $\delta > 0$ is a given constant to be used as the planner's discount rate;
iii) $U(C_t)$ is a known cardinal utility function which is continuous, concave and increasing and

$$U'(C_t) > 0, \quad U(C_t) > 0 \quad \text{for } C_t > 0$$
$$U'(C_t) \to \infty \quad \text{as } C_t \to 0$$

The problem of maximizing (3) subject to (2) can be solved using the Maximum Principle of Pontryagin, et al.[5] It is assumed that (3) converges. Attention will be directed here only to cases where steady-states occur with $C_t > 0$. Subscripts t will be dropped, except where their inclusion is instructive.

In accordance with the Maximum Principle, a necessary condition for an extremum of (3) subject to (2) to occur is that there exists a variable p_t, interpreted as the imputed demand price of a unit of unharvested resource in terms of present consumption foregone,[6] such that the current-value Hamiltonian

(4) $$H(C_t, N_t, p_t) = U(C_t) + p_t(\lambda N_t - \epsilon N_t^2 - C_t)$$

is maximized instantaneously by C_t, and

(5) $$\frac{dp_t}{dt} = p_t(\delta - \lambda + 2\epsilon N_t)$$

It is also required that

(6) $$\lim e^{-\delta t}p_t N_t = 0 \quad \text{as } t \to \infty \quad \text{and}$$
$$\lim e^{-\delta t}p_t \to 0 \quad \text{as } t \to \infty$$

(7) $$p = U'(C_t).$$

This condition states that a ncessary condition for an extremum is that the level of consumption be such that people are indifferent between consumption of that unit now, or leaving the resource unharvested to grow (investment) for future consumption. By virtue of (7) and the properties of $U(C_t)$, for any known value of p_t there is a unique C_t. (See Figure 2.) Write $C_t = C(p_t)$.

Steady State Solutions:

Let us examine the system of differential equations.

(1) $$\frac{dN}{dt} = \lambda N - \epsilon N^2 - C(p)$$

[5] See L. S. Pontryagin or M. Hestenes.

[6] One can pose the maximization problem as:

$$\max J\left[N, \frac{dN}{dt}, t\right] = \int_0^\infty U\left(\lambda N - \epsilon N^2 - \frac{dN}{dt}\right)e^{-\delta t}dt$$

This is a problem of the Calculus of Variations.

The Euler necessary condition for an extremal is:

$$U'\cdot(\lambda - 2\epsilon N - \delta) + U''\cdot\frac{dC}{dt} = 0$$

This condition is identical to equation (5) above where $U'(C) = p$ as in equation (7).

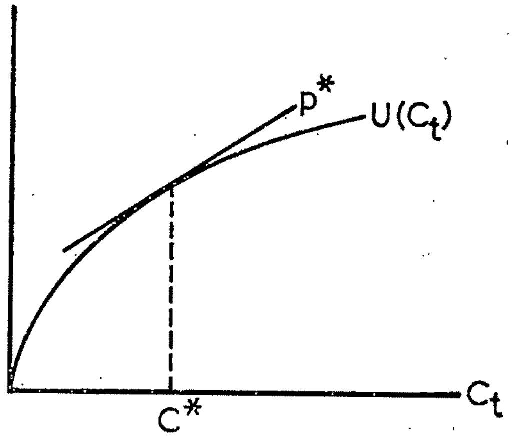


FIGURE 2

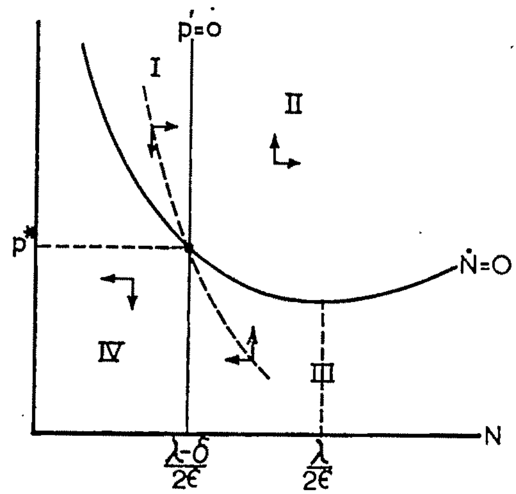


FIGURE 3

(2) $$\frac{dp}{dt} = p(\delta - \lambda + 2\epsilon N)$$

for singularities. Since $U'(C)>0$ then $p>0$ and $dp/d_t=0$ if, and only if, $N=(\lambda-\delta)/2\epsilon$. Let us restrict attention to the case where $\lambda>\delta$. In all other cases, including the non-replenishable resource one, there is no steady state solution which does not wipe out the resource. In equilibrium

$$N^* = \frac{\lambda - \delta}{2\epsilon}$$

$$C^* = \lambda N^* - \epsilon N^{*2} = \frac{\lambda^2 - \delta^2}{4\epsilon}$$

$$p^* = U'(C^*)$$

Call N^* the steady-state optimal population, and C^* the steady-state optimal consumption program.[7]

Consider next directions of movements of points in phase space. These are represented in Figure 3. Arrows to the right in Figure 3 indicate $dN/dt>0$, arrows upward indicate $dp/dt>0$. Regions are labelled with Roman numerals for identification. Briefly, along the curve $dN/dt=0$ it follows that

$$\frac{dp}{dN} = \frac{\lambda - 2\epsilon N}{C'(p)}$$

As p increases, $C(p)$ decreases and so $C'(p)$ is negative. Also, $\lambda-2\epsilon N$ is positive, zero, and negative as N is less than, equal to, or greater than $\lambda/2\epsilon$, respectively. So

$$\frac{dp}{dN} < 0 \quad \text{for } N < \frac{\lambda}{2\epsilon}$$

For fixed N, $\lambda N-\epsilon N^2-C(p)$ is an increasing function of p so below the line $dN_t/dt=0$ it follows that $dN/dt<0$ and above it $dN/dt>0$.

$$\frac{1}{p}\frac{dp}{dt}=\delta-\lambda+2\epsilon N \quad \text{so} \quad \begin{cases} \dfrac{dp}{dt}>0 & \text{for } N>\dfrac{\lambda-\delta}{2\epsilon} \\ \dfrac{dp}{dt}<0 & \text{for } N<\dfrac{\lambda-\delta}{2\epsilon} \end{cases}$$

Following Arrow (p. 18) define

$$H^0(N, p) = \max_{C^*} H(N, C, p)$$

Since H^0 is concave in N_t for given p_t,[8] and since (N_t,p_t) converges to (N^*,p^*) where $p^*\geqq 0$, the path traced by (N_t,p_t) satisfying the Pontryagin necessary conditions is optimal.

It can be shown that there exists an opti-

[7] For $N_0=(\lambda-\delta)/2\epsilon$ the consumption program $C^*=(\lambda^2-\delta c^2)/4\epsilon$ can be maintained in perpetuity.

[8] For given $p_t=\bar{p},\bar{p}=U'(C_t)$ fixes $C_t=\bar{C}$

$$H^0 = U(\bar{C}) + \bar{p}(\lambda N - \epsilon N^2 - \bar{C})$$

is concave in N.

mal path, and the optimal trajectory (marked with dashes in Figure 3) can be approximated for policy purposes. (N^*, p^*) is a saddle-point.

The Maximum Sustained Yield Consumption Program

If initially $N_0 = \lambda/2\epsilon$ then the associated consumption program $C_t = \lambda^2/4\epsilon$ can be maintained forever, and is the maximum of all such sustained programs (with appropriate initial values).[9] It has been shown, as in Crutchfield and Zellner, that this maximum sustained yield program will not maximize discounted social product for $\delta > 0$. It is shown above that such a program is nonoptimal for $\delta > 0$. That is, the steady-state *optimal* resource mass is $N_t = N^* = (\lambda - \delta)/2\epsilon$, which is less than the maximum sustained yield resource size $N_t = \lambda/2\epsilon$. And so when $N_t = \lambda/2\epsilon$ the optimal strategy from a social viewpoint is to consume more than $\hat{C}_t$.

The issue is exactly identical to the nonoptimality of Golden Rules.[10] Maximum sustained yield programs are golden rules, and steady-state optimal programs are modified golden rules.

For any given initial value of N_t a planner can assign quotas, or prices, which will direct the economy along an optimal trajectory towards a steady-state optimal consumption program.

Conclusion

When production does not involve variable inputs other than the natural resource itself, it is shown here that maximum sustained yield programs are nonoptimal. It can be shown in a much more elaborate model, that when production costs are introduced the result is to increase the optimal steady-state resource population.[11] In fact, by coincidence, a maximum sustained yield program can be optimal.

In the above simplified model, if $\delta \geqq \lambda$ the resource would be wiped out. With production costs this is not necessarily the case. The situation is as one would expect due to increasing marginal costs of producing resource product as the resource mass decreases. In fact, a more probable situation is for the resource to go unharvested at all when production costs are high.

REFERENCES

K. Arrow, "Applications of Control Theory to Optimal Growth," Institute for Mathematical Studies in the Social Sciences, Stanford Univ. 1967.

D. Cass, "Optimum Growth in an Aggregative Model of Capital Accumulation," *Rev. Econ. Stud.*, July 1965, *32*, 233–40.

J. Crutchfield and A. Zellner, "Economic Aspects of the Pacific Halibut Fishery," *Fishery Ind. Res.*, *1*, No. 1, Department of the Interior, Washington, April 1962.

G. F. Gause, *The Struggle for Existence*, Baltimore, 1934.

———, Verifications Experimentales de "Luthérie Mathématique de la Lutte Pour La Vie," *Exposes de Biometrie*, Vol. IX, 1935.

H. S. Gordon, "The Economic Theory of a Common-Property Resource: The Fishery," *J. Polit. Econ.*, April 1954, *62*, 124–42.

G. K. Goundry, "Forest Management and the Theory of Capital," *Can. J. Polit. Econ.*, Aug. 1960, *26*, 439–51.

M. Hestenes, *Calculus of Variations and Optimal Control Theory*, New York, 1966.

A. J. Lotka, *Elements of Mathematical Biology*, Dover 1956.

O. L. Mangasarian, "Sufficient Conditions for the Optimal Control of Non-linear Systems,"

$$\frac{dN}{dt} = \lambda N - \epsilon N^2 - C_t$$

For $N = \lambda/2\epsilon$ and $dN/dt = 0$ (no net change in N) the consumption path is

$$\hat{C}_t = \lambda\left(\frac{\lambda}{2\epsilon}\right) - \epsilon\left(\frac{\lambda}{2\epsilon}\right)^2 = \frac{\lambda^2}{4\epsilon}$$

The value of $N = \lambda/2\epsilon$ maximizes $\lambda N - \epsilon N^2$.

[10] See E. Phelps and D. Cass.

[11] See for example Smith and Quirk (pp. 6–24). Using a model like theirs I have gotten the result

$$N^* = \frac{\lambda - \delta}{2\epsilon} + \frac{q_2 - p}{2p\epsilon} G_2$$

where q_2 represents marginal utility of resource product, p represents the imputed price of a unit of unharvested resource and G_2 is the marginal product of the natural resource. Here $(q_2 - p)$ is nonnegative.

This is a generalization of the present model. In the present model $q_2 = p$ in the absence of production costs.

SIAM J. on Control, 1966, *4*, 139–52.
E. S. Phelps, "The Golden Rule of Accumulation: A Fable for Growthmen," *Amer. Econ. Rev.*, Sept. 1961, *51*, 638–43.
L. S. Pontryagin, et al. (translation), *The Mathematical Theory of Optimal Processes*, 1962.
J. Quirk and V. Smith, "Dynamic Economic Models of Fishing" Research Paper 22, Univ. Kansas, June 1969.
A. Scott, "The Fishery: The Objectives of Sole Ownership," *J. Polit. Econ.*, Apr. 1955, *63*, 116–24.
V. Smith, "Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.
———, "On Models of Commercial Fishing," *J. Polit. Econ.*, Mar./Apr. 1969, *77*, 181–98.
V. Volterra, *Theorie Mathématique de la Lutte Pour la Vie*, 1931.

IN MEMORIAM

PAUL THOMAS HOMAN

1893–1969

Paul T. Homan, professor emeritus of economics of the University of California at Los Angeles, and managing editor of the *American Economic Review* (1941–52), died in Washington, D.C., on July 3, 1969. At the time he was engaged in research at the Brookings Institution.

Professor Homan was born in Indianola, Iowa, on April 12, 1893. In 1914 he took his A.B. at Willamette College, where his father, the late Fletcher P. Homan, was president. During 1914–16 and 1919–20, he held a Rhodes Scholarship at Oxford, where he received an A.B. in 1919. In 1926 he obtained his Ph.D. at the Brookings Graduate School of Economics and Government.

Professor Homan began his academic career as instructor in economics at Washington University (1923–25). During 1926–27 he was assistant professor at the University of California at Berkeley. During the next two years he served in this rank at Cornell. In 1929 he became full professor, succeeding Herbert J. Davenport in the theory field. He resigned from Cornell in 1947. In 1950 he returned to academic life as professor at UCLA, where he retired in 1959. During 1959–1963 he completed his academic career at Southern Methodist University.

In the first World War, Professor Homan was attached to the British Expeditionary Forces in Mesopotamia (1916–17), and then became Second Lieutenant in the AEF during 1918–19.

Reflecting his life-long interest in public affairs, Professor Homan spent three years at Brookings in the 1930's, followed by service with wartime agencies during 1941–47, and then with the Council of Economic Advisers from 1947 to 1950.

Professor Homan's first book, *Contemporary Economic Thought* (1928), reviewed the ideas of a number of leading theorists, ranging from Marshall to Veblen. Conceding at the outset that he had been brought up "in the faith of Marshall," he insisted that he was committed "to the dogma of no one." Homan was not a system-builder but rather was a pragmatist whose abiding interest was the application of economics to the problems of public policy. Too wise to yield to the anti-theoretical currents of the time, as his essay on institutional economics in the first *Encyclopedia of the Social Sciences* clearly reveals, his bent nonetheless was toward the practical uses of economic reasoning, particularly in the service of govenment. If he had a model, it would have been John Maurice Clark or perhaps Alfred Marshall.

Accordingly, these earlier studies were followed by a series of four books of which he was coauthor, dealing with topics such as the National Recovery Administration, the Puerto Rican sugar economy, and the role of government in economic life. At the time of his death, he was engaged with Wallace Lovejoy in completing a series of studies of the petroleum industry.

Professor Homan's writings are distinguished by a remarkably graceful style through which he conveyed with exceptional clarity many subtle and complex ideas. These qualities served him well as managing editor of the *Review*. Many contributors during his tenure were the beneficiaries of his rigorous editorial standards.

In formal lectures, he put forward his views with diffidence and a measure of reserve, but in informal conversation he displayed the full reach of his lively imagination and his large knowledge of the literature of economics. Those who knew him as a friend will recall the unfailing kindness and generosity that were typical of his warm personality.

ERRATA

The *Review*, in its December 1969 issue, incorrectly printed a number of equations in R. Dorfman's article, "An Economic Interpretation of Optimal Control Theory."

The correct equations are:

p. 818

$$(1) \quad k = \frac{dk}{dt} = f(k, x, t)$$

p. 820

$$k(t + \Delta) = k(t) + k\Delta$$

$$\dot{k} = f(k, x_t, t)$$

p. 822

$$H^* = u(k, x, t) + \frac{d}{dt}\lambda k$$

$$= u(k, x, t) + \lambda k + \dot{\lambda} k$$

$$(\mathrm{I}) \quad \dot{k} = f(k, x, t)$$

p. 823

$$(\mathrm{I}') \quad \frac{\delta H}{\delta \lambda} = \dot{k}$$

p. 824

$$k = \frac{d}{dt}\frac{K}{N} = \frac{K}{N}\left(\frac{\dot{K}}{N} - \frac{\dot{N}}{N}\right)$$

$$= k\left(\frac{\dot{K}}{Nk} - n\right)$$

NOTES

The American Telephone and Telegraph Company is sponsoring post-doctoral grants to encourage independent research on the economics of public utilities. Applications are judged by an Academic Review Board composed of Professors William J. Baumol, Princeton University; Otto Eckstein, Harvard University; and Alfred E. Kahn, Cornell University. Mr. C. A. Ulffers, Jr. serves as non-voting Secretary of the Board.

Copies of the terms and conditions of the grants and application forms may be obtained by writing: Mr. C. A. Ulffers, Jr., Assistant Vice President, 195 Broadway, Rm. 2233, New York, N.Y. 10007.

New Journal: *The Intermountain Economic Review*

The *IMR* will be published semiannually at the University of Utah beginning October 1970. Leading departments of economics and agricultural economics in the intermountain West are sponsoring the *IMR*. Graduate students will be the editors, and student contributions will comprise the majority of the content. Faculty contributions will be solicited. For further information, write: Editor, *Intermountain Economic Review*, University of Utah, Salt Lake City, Utah 84112.

New Journal: *The Review of Black Political Economy*

First publication is spring, 1970 and twice a year, thereafter. Contents will include articles dealing with black economic development, book reviews, news of the Caucus of Black Economists, and information on significant economic activities in the black community. The interim editor is Professor Robert S. Browne, Fairleigh Dickinson University. The cost is $2.00 per issue; $5.00 per year. Orders may be placed with the publisher: The Black Economic Research Center, 112 W. 120th St., New York, N.Y. 10027.

International Comparative Economics: New Programme in Vienna, Austria

The Austrian Institute of Economic Research (Vienna) has established a department for International Comparative Economics. The emphasis will be on comparative analyses of general tendencies in economic development in eastern and western countries. Research will at first be limited to Austria and a few smaller industrialized countries in the West on the one hand and to her socialist neighbors, Czechoslovakia, Hungary, and Yugoslavia, on the other. Specific subjects, such as investment decisions, labor market problems, industrial efficiency, incentive schemes, foreign trade, will be studied primarily with quantitative-empirical methods by competent economists from both eastern and western countries, working in close collaboration.

Economists from eastern countries will as a rule be on assignment or leave from their home institutions for periods of twelve months or less. Economists from western countries who are interested in the proposed area of research and, preferably, have experience in comparative-systems analysis are invited to participate in this program. The Austrian Institute of Economic Research will make its facilities (office space, library, secretaries) available to visiting research fellows without charge, but candidates will have to secure their own funds to finance living expenses in Vienna. Depending on the chosen research project, the length of residence in Vienna should be up to twelve months.

Inquiries and applications may be addressed to Professor Dr. Franz Nemschak, Director, Austrian Institute of Economic Research, Postfach 91, 1103 Vienna, Austria.

The Asia Foundation has provided the American Economic Association with a fund to be used to assist students and visiting scholars from Asia to attend the annual meeting of the American Economic Association. The maximum amount available to an individual applicant is $100. Inquiries should be addressed to the AEA, 629 Noyes Street, Evanston, Illinois 60201.

Clearinghouse For Refugee Economists

Events in Poland and Czechoslovakia in the last few years have created a new group of refugee professionals, among them many economists. To help these people find positions that would use their professional training and experience as economists, the Association for the Study of Soviet-type Economies (*ASTE*) and the Association for Comparative Economics have undertaken to act as a clearinghouse.

Institutions which may be able to offer teaching or research positions to these persons and persons interested in such positions are invited to get in touch with: Robert W. Campbell, Executive Secretary, *ASTE*, Department of Economics, Indiana University, Bloomington, Indiana 47401.

Miss Agnes R. Miller, 5 Gracie Square, New York, N. Y. 10028, will donate her collection of 34 *TNEC* monographs to any interested person or organization.

The Committee on International Exchange of Persons announces availability of economist Aurelio Macchioro (Italy) nominated for Fulbright-Hays travel grant, seeking 1970–71 lecturing or research appointment in U.S. University or college.

Inquiries should be addressed to: Miss Grace E. L. Haskins, Program Officer, Conference Board of Associated Research Councils, 2101 Constitution Avenue, Washington, D.C. 20418 or via telephone—Area Code 202, 961–1948.

The 1970 Western Resources Conference will be held at the University of Denver on July 8, 9, and 10. The theme will be Urban Demands on Natural Resources, considered under five topics; Air, Water, Raw Materials, Land, and Choices. Over twenty speakers from the Rocky Mountain region and other parts of the United States will appear in the five sessions. John J. Schanz, Jr. professor of natural resources at the University of Denver is chairman for the 1970 conference.

The National Institute of Social and Behavioral Sciences will hold its regular sessions for contributed papers at the 137th annual meeting of the American Association for the Advancement of Science, December 26–31, 1970, in Chicago. Economists interested in presenting a paper at these sessions are invited to forward titles and abstracts of 300 words to Donald P. Ray, Director, National Institute of Social and Behavioral Science, 863 Benjamin Franklin Station, Washington, D.C. 20044 by August 20.

Deaths

Spurgeon Bell, College of Business Administration, The University of Texas, Austin, Dec. 1968.

Elmer J. Brown, College of Business and Public Administration, University of Arizona, Jan. 7, 1970.

Alonzo B. Cox, professor emeritus of marketing, College of Business Administration, The University of Texas, Austin, Dec. 1968.

Eleanor H. Grady, Jan. 6, 1970.

John Ise, professor emeritus of economics, University of Kansas, Mar. 26, 1969.

Randolph G. Kinabrew, associate dean, School of Business Administration; professor of economics, University of Mississippi.

Max F. Millikan, director, Center for International Studies, Massachusetts Institute of Technology, Dec. 14, 1969.

Ernest M. Patterson, professor emeritus, University of Pennsylvania, and president of the American Academy of Political and Social Science, Nov. 9, 1969.

William J. Phillips, May 29, 1969.

Ella J. Polinsky, Sept. 22, 1969.

V. K. Ramaswami, New Delhi, Oct. 18, 1969.

Ernest Sheppard, Purdue University, Jan. 5, 1970.

Walter E. Spahr, professor emeritus of economics, School of Commerce, New York University, Jan. 19, 1970.

Helen Wright, dean emeritus, University of Chicago, Aug. 14, 1969.

Retirements

John M. Cassels, department of economics, University of Colorado, June 1970.

Raymond W. Coleman, dean, emeritus professor of economics and management, College of Business Administration, University of Illinois at Chicago Circle, Aug. 31, 1969.

Jessamon Dawe, College of Business Administration The University of Texas, Austin, June 1969.

Howard L. Hoag, Purdue University.

George H. Newlove, professor emeritus, College of Business Administration, The University of Texas, Austin, Dec. 1968.

John P. Troxell, professor of economics, University of Mississippi.

Colston E. Warne, professor of economics, Amherst College, June 1970.

Visiting Foreign Scholars

Henryk Flakierski, Central School of Planning and Statistics, Warsaw, Poland: associate professor, McGill University.

S. Herbert Frankel, Oxford University: professor of economics, University of Virginia, first semester 1970.

Admantios Pepelasis, Athens, Greece: visiting professor of economics, Virginia Polytechnic Institute.

Samuel B. Saul, University of Edinburgh: Ford Foreign Faculty Fellowship, Stanford University, 1969–70.

Promotions

John W. Allen: associate professor of economics, Texas A&M University.

Manuel Avila: associate professor of economics, Villanova University.

Lee D. Badgett: assistant professor of economics, US Air Force Academy.

Alexander Belinfante: assistant professor of economics, School of Commerce, New York University.

Keith Brown: associate professor of economics, Purdue University.

J. Martin Carovano: associate professor, Hamilton College, 1969–70.

Lawrence L. Crum: professor of finance, College of Business Administration, The University of Texas, Austin, 1969–70.

Eugene J. Devine: associate professor, department of economics and commerce, Simon Fraser University.

James J. Diamond: professor, department of economics, De Paul University, Jan. 1970.

Arnold A. Faden: associate professor of economics, Iowa State University, Jan. 1970.

Curtis Gilroy: assistant professor of economics, Wells College.

William I. Greenwald: professor of economics, The City College, Uptown University, Jan. 1970.

Charles W. Hackett, Jr.: associate professor of finance, College of Business Administration, The University of Texas, Austin, 1969–70.

Abraham Hirsch: professor, department of economics, Brooklyn College.

William T. Hold: associate professor of finance, College of Business Administration, The University of Texas, Austin, 1969–70.

Richard A. Holmes: professor, department of economics and commerce, Simon Fraser University.

James Hulbert: assistant professor, Graduate School of Business, Columbia University, Jan. 1970.

Joseph Humphrey: assistant professor, department of economics, Occidental College.

Clinton E. Jencks: professor, San Diego State College, Sept. 1970.

Mahmood H. Khan: associate professor, department of economics and commerce, Simon Fraser University.

Lennis M. Knighton: associate professor of accounting, College of Business Administration, The University of Texas, Austin, 1969–70.

Mark L. Ladenson: assistant professor of economics, Michigan State University.

Kermit C. Larson: associate professor of accounting, College of Business Administration, The University of Texas, Austin, 1969–70.

Robert E. Lucas: professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University.

Bogdan Mieczkowski: professor of economics, Ithaca College.

John M. Munro: associate professor, department of economics and commerce, Simon Fraser University.

John A. Nehring: assistant professor of economics, US Air Force Academy.

Nilan Norris: professor of economics, Herbert H. Lehman College, City University of New York.

Bernard Okun: professor of economics, Brooklyn College, City University of New York.

C. G. Plourde: assistant professor of economics, The University of Western Ontario, Jan. 1970.

Joel Popkin, Price Research Division: assistant commissioner, Bureau of Labor Statistics for Prices and Living Conditions.

Russell G. Pounds: assistant professor of economics, Iowa State University, Feb. 1970.

Gordon Pye: professor, department of economics, University of California, Berkeley, Nov. 1969.

Leonard A. Rapping: professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University.

Richard Roll: associate professor of economics and industrial administration, Graduate School of Industrial Administration, Carnegie-Mellon University.

Allan B. Rosenberg: associate professor of economics, West Liberty State College.

Nathan Schmukler: professor of economics, Brooklyn College, Jan. 1970.

Jerome E. Schnee: assistant professor, Graduate School of Business, Columbia University, July 1969.

Francis H. Schott: vice president and economist, Equitable Life Assurance Society of the U.S.

Mark B. Schupack: professor of economics, Brown University.

Richard Simpson: associate professor of accounting, University of Massachusetts, Feb. 1970.

Rohini P. Sinha: associate professor of economics, Muhlenberg College.

Joseph Taffet: associate professor of economics, The City College, Uptown University, Jan. 1970.

Yiannis P. Venieris: associate professor, San Diego State College, Sept. 1970.

Gordon Wagner: assistant professor of economics, Wells College.

Maurice D. Weinrobe: assistant professor of economics, Michigan State University.

Richard E. Wilson: associate professor of economics, University of Nevada.

Administrative Appointments

Robert E. Barckley: chairman, department of economics, San Diego State College, Feb. 1970.

John J. Casson, Research Institute Investors Service: senior economist, Savings Banks Association of New York State.

Paul G. Craig: vice president, Academic Affairs, Florida State University.

Daniel E. Diamond: acting dean, School of Commerce, New York University, Feb. 1970.

John P. Herzog: acting chairman, department of economics and commerce, Simon Fraser University.

Juanita M. Kreps: dean, Woman's College and assistant provost, Duke Univeristy.

R. F. Lanzillotti: dean, College of Business Administration, University of Florida.

Hal B. Lary: vice president, research, National Bureau of Economic Research.

Richard H. Laube: dean, College of Business Administration, Valparaiso University, Jan. 1970.

Robert E. Lipsey: vice president, research, National Bureau of Economic Research, July 1970.

Kenneth D. McIntosh: director, Division of Resource Management, West Virginia University.

Ronald I. McKinnon: chairman, department of economics, Stanford University.

F. Ray Marshall: chairman, department of economics, The University of Texas, Austin, fall 1970.

Joseph R. Mason: acting chairman, department of economics, State University of New York, Brockport.

Richard U. Miller: associate director, Industrial Relations Research Institute, University of Wisconsin.

Joseph P. Mooney: professor of economics, chairman of department, La Salle College.

Leon N. Moses: chairman, department of economics, Northwestern University.

Kenneth L. Parkhurst: chairman, department of economics, John Carroll University, July 1969.

Gerald G. Somers: chairman, department of economics, University of Wisconsin.

James H. Stauss, Grinnell College: executive vice president, provost, and professor of economics, The Colorado College.

James L. Stern: director, Industrial Relations Research Institute, University of Wisconsin.

Clifton R. Wharton: president, Michigan State University.

Edgar W. Wood: acting chairman, department of economics and finance, University of Mississippi.

Joseph R. Zandarski: chairman, departments of business and economics, University of Scranton.

New Appointments

Marcus Alexis, Northwestern University: visiting professor, University of California, Berkeley, fall 1970.

Saul Arbess, University of Saskatchewan: visiting

assistant professor, department of economics and commerce, Simon Fraser University.

Sven W. Arndt, University of California, Los Angeles: visiting assistant professor of economics, Crown College, University of California, Santa Cruz.

Ron N. Bagley: assistant professor of accounting, College of Business Administration, The University of Texas, Austin.

John R. Baldwin: assistant professor, department of economics, Queen's University.

Jon H. Barrett: assistant professor, department of management; joint appointment, department of psychology, College of Business Administration, The University of Texas, Austin, Jan. 1970.

R. N. Batra: assistant professor, department of economics, The University of Western Ontario, July 1970.

George W. Bogdanow: instructor, department of economics and commerce, Simon Fraser University.

Peter Bohmer, Massachusetts Institute Technology: San Diego State College, Sept. 1970.

John P. Bonin: lecturer, department of economics, Wesleyan University, 1970–71.

John O. Bornhofen: associate professor of money and banking, Kalamazoo College.

R. S. Boyer: lecturer, department of economics, The University of Western Ontario, July 1970.

David W. Breneman: assistant professor, department of economics, Amherst College, July 1970.

Basil Browne, University of Idaho: assistant professor of economics, Allegheny College.

Victor P. Buell: associate professor of marketing, School of Business Administration, University of Massachusetts, Feb. 1970.

Walter R. Butcher, Washington State University: National Water Commission.

James R. Cantwell, Washington University: assistant professor, department of economics, Texas A&M University.

David Cass, Yale University: professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University.

Albert M. Chammah: assistant professor, department of management; joint appointment, department of psychology, College of Business Administration, The University of Texas, Austin.

Chu-yuan Cheng: associate professor, department of economics, Lawrence University, Sept. 1970.

Gregory C. Chow, IBM: professor of economics and director, Econometric Research Program, Princeton University.

Robert E. Coeberd, Jr.: director of public finance analysis, American Enterprise Institute for Public Policy Research.

Craig G. Coelen: instructor, department of economics, Northeastern University, 1970–71.

Barry E. Cushing: assistant professor of accounting, College of Business Administration, The University of Texas, Austin, Jan. 1969.

Horace B. Davis: professor of economics, Hofstra University.

J. E. Dittrich: lecturer, management department, University of Nevada, 1969–70.

Clifford Dobitz: instructor of economics, North Dakota State University.

Michael A. Duggan: visiting associate professor of business law, department of general business, College of Business Administration, The University of Texas, Austin, Sept. 1969.

Elizabeth Durbin: lecturer in economics, School of Commerce, New York University, Feb. 1970.

Dean S. Dutton, Brigham Young University: visiting assistant professor, department of economics, Texas A&M University, 1969–70.

Robert Edmister, Ohio State University: assistant professor of management, Purdue University.

Douglas M. Egan, Oregon State University: visiting associate professor, department of economics and commerce, Simon Fraser University.

Karl Egge: instructor in economics, Macalester College, fall 1970.

Gano S. Evans, Northern Arizona University: associate professor, management department, University of Nevada, 1969–70.

George A. Eversull, San Jose State College: lecturer, accounting department, University of Nevada, 1969–70.

Albert Fishlow: professor of economics, Stanford University.

Terry Glover, Purdue University: Ohio State University, Apr. 1970.

Gerald S. Goldstein: assistant professor of economics and urban affairs, Northwestern University, fall 1970.

David H. Greenberg: economics department, The RAND Corporation, Aug. 1969.

Robert L. Greene, University of Georgia: assistant professor of economics, Auburn University, fall 1970.

Hans M. Gregersen: assistant professor, joint appointment, department of agricultural economics and school of forestry, University of Minnesota.

David M. Grether: associate professor of economics, California Institute of Technology.

Walter Haessel: lecturer, department of economics, The University of Western Ontario, July 1970.

Alfred J. Hagan: assistant professor of international business and resources, department of marketing, College of Business Administration, The University of Texas, Austin.

Kenneth N. Hamilton: instructor of economics, School of Business, West Liberty State College.

Jon P. Harkness: assistant professor of economics, Northwestern University, fall 1970.

John C. Hause: research fellow, National Bureau of Economic Research, 1970 calendar year.

Oli Hawrylyshyn: assistant professor, department of economics, Queen's University.

Lawrence A. Herbst: lecturer, department of economics, Vassar College, 1970–71.

Stephen J. Hiemstra: assistant to the administrator, Food and Nutrition Service, U.S. Department of Agriculture.

Werner Hildenbrand: visiting professor of economics, Stanford University.

Gary L. Holstrum: assistant professor of accounting, College of Business Administration, The University of Texas, Austin, July 1969.

Charles W. Howe: professor of economics, University of Colorado, Sept. 1970.

Jay M. Hughes: professor, joint appointment, department of agricultural economics and school of forestry, University of Minnesota.

Stanley R. Johnson, University of Missouri-Columbia: visiting associate professor, University of California, Davis.

J. M. Jondrow: lecturer, department of economics, The University of Western Ontario, July 1970.

Eric L. Jones: professor of economics, Northwestern University, fall 1970.

James E. Jones, Jr.: visiting professor of industrial relations and law, University of Wisconsin.

Pamela H. Kacser, American University: assistant chief, division of economic studies, Bureau of Labor Statistics, and associate professor, doctoral program of policy sciences, University of Buffalo.

Robert R. Keller, The University of Wisconsin: assistant professor, department of economics, California State College, Long Beach, fall 1970.

John W. Kendrick, George Washington University: visiting professor, department of economics and commerce, Simon Fraser University.

Darwin D. Klingman: assistant professor of business statistics, operations research, department of general business, College of Business Administration, The University of Texas, Austin, June 1969.

Kenneth E. Knight: department of management, College of Business Administration, The University of Texas, Austin, Jan. 1969.

L. Emil Kreider, Southern Illinois University: chief of Special Mission, AID, Paraguay, 1970.

Michael Lavelle: assistant professor of economics, John Carroll University, Sept. 1969.

Michael Landsberger: research staff associate, National Bureau of Economic Research, June 1970.

William N. Layher, University of Wisconsin: lecturer, economics department, University of Maryland, Sept. 1969–70.

John O. Ledyard: assistant professor of economics, Northwestern University, fall 1970.

Solomon B. Levine, University of Illinois: professor of business, economics, and industrial relations, Industrial Relations Research Institute, University of Wisconsin.

Jung Chao Liu, McGill University: associate professor of economics, State University of New York, Binghamton, Sept. 1970.

Kenneth D. Loeffler, Monmouth College: associate professor, finance department, University of Nevada, 1969–70.

Michael C. Lovell: professor of economics, Wesleyan University, Nov. 1969.

John R. K. McGuire: assistant professor, department of economics, University of Manitoba.

Joseph A. McKinney, Michigan State University: assistant professor, University of Virginia, Feb. 1970.

Marshall E. McMahon: instructor in economics, Gettysburg College, 1970–71.

James W. Martin, University of Kentucky: professor of political science, Graduate School, University of Louisville, Jan. 1970.

Robert L. C. Miller, University of Oregon: lecturer, accounting department, University of Nevada, 1969–70.

Alan S. Milward: associate professor of economics, Stanford University.

Albert A. Montgomery: department of agricultural economics, Washington State University.

Lawrence H. Officer, Harvard University: professor of economics, Michigan State University, Sept. 1970.

Gordon W. Paul: associate professor of marketing, School of Business Administration, University of Massachusetts, Sept. 1969.

Philip Perry: instructor, department of economics, Occidental College, July 1970.

Lee E. Preston, University of California, Berkeley: professor of American enterprise, School of Management, State University of New York at Buffalo, July 1969.

Henry O. Pruden: assistant professor, department of marketing, College of Business Administration, The University of Texas, Austin.

V. S. Rao: lecturer, department of economics, The University of Western Ontario, July 1970.

Henry Rempel: assistant professor, department of economics, University of Manitoba.

Robert F. Rooney: associate professor, department of economics, California State College, Long Beach, spring 1970.

Daniel H. Saks, Princeton University: instructor of economics, Michigan State University, Sept. 1970.

Emile B. Sawaya: instructor, department of economics, DePauw University, Jan. 1970.

Gene R. Schwab: assistant professor of economics, Mankato State College.

Thomas A. Sears: assistant professor, department of economics, Amherst College, July 1970.

Harold S. Sekiguchi, Drake University: associate professor, marketing department, University of Nevada, 1969–70.

Merrill K. Sharp: instructor, department of economics, Iowa State University.

Alan Shelly, University of California, Berkeley: instructor of economics, Michigan State University, Sept. 1970.

William L. Sherrill: instructor, department of economics, Old Dominion University, Feb. 1970.

Christopher A. Sims: research fellow, National Bureau of Economic Research, 1970–71.

Charles H. Smith: assistant professor of accounting, College of Business Administration, The University of Texas, Austin, Sept. 1969.

Robert F. Smith, Louisiana State University: visiting fellow, Institute for Research on Poverty, University of Wisconsin, Madison, Jan. to May 1970.

Kalburgi M. Srinivas: assistant professor, department of management, College of Business Administration, The University of Texas, Austin, July 1969.

Arnold Studenmund: instructor, department of economics, Occidental College, July 1970.

Steven M. Swanson: instructor, department of economics, Northwestern University, 1970–71.

Hirsh L. Tadman: instructor in banking, School of Commerce, New York University, Feb. 1970.

Ben B. Tarwater, Bowling Green State University: assistant professor of economics, University of Texas at El Paso.

Kenna C. Taylor: instructor of economics, School of Business, West Liberty State College.

Alfred Tella, President's Commission on Income Maintenance Programs: research professor of economics and director, Income Maintenance Project, Georgetown University.

Edward Tower, Harvard University: assistant professor of international economics, Fletcher School of Law and Diplomacy, Tufts University.

Joan R. Tron: director of publications, National Bureau of Economic Research.

T. J. Valentine: lecturer, department of economics, The University of Western Ontario, July 1970.

Roy Vogt: assistant professor, department of economics, University of Manitoba, July 1970.

Herbert W. Warburton: visiting assistant professor, department of economics and commerce, Simon Fraser University.

Richard J. Ward, Bureau for Near East/South Asia, AID: manager, International Economics Consulting, Peat, Marwick, Mitchell & Co., Washington.

Rosaline H. Weaver: lecturer, accounting department, University of Nevada, 1969–70.

Charles L. Weber: instructor in economics, Ball State University.

E. Roy Weintraub: assistant professor of economics, Duke University.

William Wilkie, Stanford University: assistant professor of management, Purdue University.

Thomas Willett, Harvard University: associate professor of international economics, Fletcher School of Law and Diplomacy, Tufts University.

Robert C. Witt: assistant professor of insurance, department of finance, College of Business Administration, The University of Texas, Austin, Jan. 1970.

Paul D. Zook, Southern Methodist University: professor of economics, University of Texas at El Paso.

Leaves for Special Appointments

Irma Adelman, Northwestern University: Fellow Center for Advanced Study in the Behavioral Sciences, Palo Alto, Sept. 1970.

Stuart Altman, Brown University: The Urban Institute, Washington.

Frank P. R. Brechling, Northwestern University: professor of economics, Ford Foundation Fellowship, University of Essex, England.

Douglas R. Carmichael, The University of Texas, Austin,: manager of auditing research, American Institute of Certified Public Accountants, New York.

Calvin P. Blair, The University of Texas, Austin,: Ford Foundation, social science advisor for Mexico and Central America.

E. Jay Hall, The University of Texas, Austin,: Illinois Commission on Mental Health, spring 1970.

Richard F. Kosobud, University of Illinois at Chicago Circle: research associate, The Institute of Economic Research, Hitotsubashi University, Tokyo.

James W. Meehan, Jr., Northwestern University: economic adviser, Federal Trade Commission, Washington, 1970–71.

Dale T. Mortensen, Northwestern University: assistant professor of economics, University of Essex, England, Sept. 1970.

Edward Neuner, San Diego State College: deputy dean for Instructional Affairs, Academic Planning, Chancellor's Office, California State Colleges, Los Angeles, Feb. 1970.

Roger C. Noll, California Institute of Technology: senior fellow, The Brookings Institution, 1970–71.

Simon Rottenberg, Duke University: International Labour Office, Geneva, Switzerland.

Ryuzo Sato, Brown University: Ford Faculty Fellowship, University of England, Cambridge.

Robert E. Will, Carleton College: visiting fellow, Institute of Development Studies, University of Sussex, 1970–71.

Resignations

Richard Bergen, The University of Texas, Austin.

Laurence deRycke, Occidental College, July 1970.

S. B. Gupta, The University of Western Ontario, June 1970.

Robert E. Hall, University of California, Berkeley: Massachusetts Institute of Technology, June 1970.

Michael J. Hartley, Duke University, Aug. 1970.

Bob S. Hodges III, College of Business Administration, The University of Texas, Austin.

C. D. Hodgins, The University of Western Ontario: Dominion Bureau of Statistics, Ottawa, June 1970.

David Holland, College of Business Administration, The University of Texas, Austin.

Carlyle D. Hughes, College of Business Administration, The University of Texas, Austin: George Washington University.

Erich Klinkmuller, Saint Louis University: University of Berlin.

Charles McConnel, Occidental College, July 1970.

T. R. Robinson, The University of Western Ontario: Federal Government of Canada, Ottawa, June 1970.

Edward Shaffer, Occidental College, July 1970.

James Shepherd, Purdue University, June 1970.

Donald S. Shoup, National Bureau of Economic Research: Social Science Research Council.

Donald V. Stuchell, College of Business Administration, The University of Texas, Austin: Ohio University.

Robert E. Willard, University of Texas at El Paso, June 1970.

L. S. Zudak, The University of Western Ontario: University of Miami, June 1970.

Miscellaneous

Glen G. Cain, University of Wisconsin: associate editor, *The Journal of Human Resources*.

Robert J. Lampman, University of Wisconsin: editor, *The Journal of Human Resources*.

S. V. Wantrup, University of California, Berkeley: member, International Marine Science Affairs Panel, National Academy of Sciences.

The American Economic Review

ARTICLES

SEPTEMBER 1970

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September and December.

• Membership dues of the Association are $10.00 a year, $8.00 of each year's dues being in payment of a year's subscription both to the *Review* and the *Journal of Economic Literature*. Subscriptions by nonmembers are $10.00 a year. Single copies of the *Review*, $2.00; of the *Journal*, $2.00.

• Correspondence relating to the *Paper and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Production, Price, and Inventory Theory

By GEORGE A. HAY*

This paper is an attempt to derive empirically testable hypotheses regarding the principal determinants of firms' decisions on production, price, and finished goods inventory. The general approach to the problem is that many of the same factors which affect the optimal value for one variable will also influence decisions on the other two, and that a "proper" model must take into account the interdependence of these variables and the simultaneous nature of the decisions involving them. This is in contrast to literature on the theory of inventories (see Paul Darling and Michael Lovell) in which the firm is assumed to determine the optimal level of inventories with the rate of production taken as given and with price held constant. Similarly there are theories of price formation (see Otto Eckstein and Gary Fromm) in which the rate of production is assumed to have been determined previously, and in which the level of inventories is often ignored entirely. The present paper will attempt to present an "integrated" model of firm behavior in which decisions on all relevant variables are assumed to result from a single optimization process.

A principal distinction between this study and previous work in this area (see Gerald Childs and Charles Holt) is the inclusion of price as one of the decision variables.[1] In the past the assumption has been made that price is fixed and therefore that quantity demanded is a datum to the firm. In anything but a purely competitive model, however, the firm does exercise some control over price. The rational firm would use its current pricing policy to select the specific price-quantity combination on the demand curve that best contributes to its overall scheme of profit maximization. Thus the firm whose demand curve is not constant over time but fluctuates from period to period on a random and/or seasonal basis might view price adjustments as one means of achieving production stabilization. If this were so we might expect to find that an increase in demand would be met by continuing to produce at or near the previous rate and raising price to clear the market. More realistically, the entire kit of adjustment tools—inventory, backlog, and price—would be used in some combination to absorb all or part of the increased demand, the specific result depending not only on the particular cost structure assumed, but also on the firm's estimates of what demand will be for several periods hence. The important point is that price must certainly be included as one of these tools.

In the remainder of the paper we construct a model which includes many of the variables which are important at the individual firm level, and which treats decisions regarding those variables as being essentially interdependent. The subsequent section discusses the behavioral assumptions which underlie the model and expresses the model in mathematical form. The first-order conditions for maximizing expected profits generate a set of linear decision rules for production, price, and finished goods inventory. On the assump-

* Assistant professor of economics at Yale University. I wish to acknowledge the valuable advice of Gerald Childs at an early stage in the preparation of this paper. A referee also provided useful suggestions.

[1] Edwin Mills developed a model which included price. He was able to derive approximations to the true decision rules which held up reasonably well under empirical investigation. The present study attempts to derive an exact set of decision rules.

tion that firms do attempt to maximize expected profits, these decision rules are suitable for empirical investigation with the appropriate data. In Section II, the model is solved numerically with representative cost parameters. The resulting decision rule coefficients provide predictions regarding the actual regression equations which should be fulfilled if the model accurately reflects the working of the real world. Finally, in Section III, the regressions are performed on two industry groups, and the results compared with the predictions.

I. *The Model*

Behavioral Assumptions

The model is intended to represent a firm which chooses the levels of the variables it controls in order to maximize the expected value of discounted future profits over a time horizon. The variables involved are the rates of production and shipments, the level of finished goods inventories, the backlog of unfilled orders, and price. These variables are not independent, being related by various definitional and market constraints, including the demand equation. Each of the variables serves a particular function, and associated with each of these variables are certain costs. These functions and costs form the basic elements of the model.

A positive level of unfilled orders is, practically speaking, an unavoidable phenomenon for most firms which undertake any production to order. The timing of the arrival of new orders is not in general subject to control by the firm. The development, design, and production of each order takes time, so that there will typically be some work still in process when a new order is received. Even beyond this, however, a *higher* level of unfilled orders may be a useful alternative for the firm because it permits smoothing of production within the period and accumulation of optimal size production batches. This is particularly true when production is a multistage process and several items which are eventually individualized to the specific requirements of their respective purchasers could nonetheless go through several stages of the production process together. On the other hand, there are costs associated with a high level of unfilled orders. As the backlog gets larger and the lead time longer, there is increased danger of cancellation of orders, penalty costs for expediting particular orders, and probable loss of future sales.

This suggests that there is some positive level of unfilled orders which balances the cost savings attributable to an order backlog and the penalties associated with too large a backlog so that the net saving is maximized. We might refer to this as the "desired" level of unfilled orders in the sense that, if there were no other considerations involved, this is the level the firm would try to maintain. We will assume that the desired level of unfilled orders, U^*, is a linear function of the rate of production:

$$U_t^* = c_{13} + c_{14}X_t$$

As stated by Childs:

> Penalty costs for cancellation of orders, for expediting particular orders in response to customer requests and the probability of loss of future sales all increase as the size of the backlog increases and the lead time lengthens. However, more flexibility in production arises as backlog mounts. Therefore, as backlog and lead time increase the costs related to inflexibility of production decline. Average lead time is approximately the ratio of backlog to the rate of production. Then for every rate of production, the cost associated with varying size of backlog is the sum of monotonically rising and monotonically falling components over the relevant range and has a minimum, U_t^*, the optimal level of U_t. [p. 10][2]

[2] It is probably true that U_t^* is a function of only

Furthermore, we will assume that the cost of deviating from the desired level increases quadratically with the size of the deviation and is the same in either direction. Thus the net contribution to total cost of a given level of unfilled orders is given by:

$$c_{11} + c_1(U_t - U_t^*)^2$$

where U_t is the actual level of unfilled orders at the end of t, and c_{11} is a constant (which may be negative).

Thus we have separated all the costs and cost savings specifically associated with an order backlog into two parts; the first being the net contribution to total costs of the desired level, c_{11}, and the remainder reflecting the additional costs of deviating from the desired level, recognizing that when *all costs* which the firm incurs are considered, the optimizing level of unfilled orders may differ from the desired level. The rational firm may, for example, be willing to deviate slightly from its desired level if doing so will make it possible to avoid a substantial increase (or decrease) in the rate of production from one period to the next.

A similar argument can be used to explain the existence of a positive level of inventories. Certain cost savings accrue from the added flexibility made possible in production; on the other hand, an inadequate level of inventories can have serious consequences for future sales through loss of goodwill which may be caused by a "stock out"—a firm's inability to fill an order from a customer who requires immediate delivery. A firm not only loses out on the sales corresponding to that particular order, but may suffer the loss of future sales as well if the disappointed firm switches preferences in favor of another supplier. This effect might most conveniently be handled as a cost whose expected value is associated with an inadequate level of finished goods inventories.

We assume that the desired level of inventories is related to the quantity to be shipped during the period. This can be considered an approximation to the optimal lot size formulas in the operations research literature (see Holt et al. pp. 56–57). Specifically we assume that the relation:

$$H_t^* = c_{23} + c_{24}S_t$$

is valid over the relevant range, where H_t^* is the desired level of finished goods inventories at the end of the period t, S_t is shipments, and c_{23} and c_{24} are appropriate constants. Furthermore, we assume that the cost associated with being away from the desired level of inventories may be approximated by a quadratic such that the overall contribution of inventories to total costs is represented by:

$$c_{21} + c_2(H_t - H_t^*)^2$$

where c_{21} reflects the net effect of the "desired" level of finished goods inventories.

Unit costs of production within a time period are assumed constant. Beyond this, however, are costs associated with changes in the rate of production from period to period which are independent of the actual level of production. These include various setting up costs and costs of hiring and firing where a change in the work force is required. We express these costs simply as:

$$c_3(X_t - X_{t-1})^2$$

This captures the notion that changes are costly in either direction and that large changes are likely to be relatively more

that part of X_t which is production-to-order. However, introducing production-to-order and production-to-stock as separate variables complicates the analysis and furthermore requires an arbitrary aggregation procedure at the end since the only observed variable is total production. In addition, the nature of the results not is affected so long as the cost of changing the rate of production (discussed below) is independent of the mix between order and stock. (Similar remarks are applicable to the inventory decision.)

costly over a certain range than small ones. The general idea is that firms do seem to attach considerable weight to the stability of the rate of production and the size of the work force. This has the effect of making the previous period's output a factor of considerable importance in the determination of the output for the current period.

Demand for the product of the firm (in the sense of the entire demand curve) is not constant over time but is assumed to shift from period to period, in response partly to random factors, and partly to factors which the firm observes but cannot control (e.g., the level of national income). To make the model amenable to analytical treatment we must give a specific form for the demand curve. In this model, therefore, we assume that demand is of the form:

$$O_t = Q_t - bP_t$$

when O_t is new orders and P_t is price in period t. This is a linear demand curve of constant slope b which shifts over time in parallel movements, the extent of the shift determined by the quantity intercept term Q_t. For reasons discussed below, this demand curve is the schedule of the quantity demanded from the firm at various prices when all firms charge the same price.

This leads to the final cost to be considered, viz., that associated with changes in price. It may not be common to think about costs of changing price yet we observe that firms are generally reluctant to do so in terms both of frequency and magnitude. There are, of course, certain out-of-pocket costs which must be met—the necessity of revising price books and possibly some additional advertising expense to announce the new price.

Probably a more important influence on firm behavior, however, are those costs which are never actually observed but may be thought of more as a type of opportunity cost, the amount of money a firm would be willing to pay, *ceteris paribus*, to avoid changing price. This reluctance is due generally to the risk associated with imperfect knowledge about reaction by competitors to a price change. We cite William Fellner for a description of the reasoning involved:

> Each firm knows that others have different appraisals [of the appropriate policy for the maximization of industry profits] and that they are *mutually* ignorant of what precisely the rival's appraisal is. Consequently, no firm can be *sure* whether the move of a rival is toward a profit-maximizing quasi-agreement or towards aggressive competition; and no firm can be sure how its own move will be interpreted. This is where the desire to avoid aggressive competition enters as a qualifying factor. Even where leadership exists, the leader's moves may be misinterpreted as aiming at a change in relative positions rather than as being undertaken in accordance with the quasi-agreement. [p. 179]

There are, of course, other arguments which might be offered for the inclusion in the criterion function of a penalty for price changes. The important consideration is that the firm acts "as if" there were costs attached to changes in price, whether these are explicit, out-of-pocket costs, or more of an opportunity-cost, implicit type which arises from uncertainty about rivals' reactions to its own price decisions.

We might assume initially that the cost of changing price can be expressed as:

$$c_4(P_t - P_{t-1})^2$$

The symmetry assumption regarding the cost of changing price is perhaps bothersome but extremely convenient mathematically. If we regard the demand curve as based on the assumption that other firms will always imitate price changes by our firm, then the quadratic penalty can be taken to represent the fear that for a price increase, the firm will not be followed,

and on a price decrease, the firm will be undercut.

There may be certain circumstances, however, under which the reluctance to change price might be considerably diminished. If the incentive for a firm to initiate a price increase is the result of an increase in labor or raw materials cost (which presumably would affect all firms in the industry to a similar, if not identical, degree), it is likely that such a move would be welcomed by rivals as an opportunity for them to restore the margin which existed prior to the increase in direct costs, and hence the price increase would be matched. Indeed, failure to do so by any single firm might well be interpreted as aggressive behavior by rivals in the same sense as a price cut in the absence of any changes in cost. To the extent, then, that firms in the industry follow this type of increase completely, the cost associated with initiating such a change is likely to be insignificant. (A similar argument can be given for price cuts which are occasioned by a drop in direct costs.)

Therefore, we can amend the model to assume that the firm is sensitive to changes in markup rather than changes in the absolute level of price. Cost-induced price changes can be initiated without fear of not being followed and hence the firm does not attribute a penalty to such a move. Changes in price which are not related to direct costs, or failure to change price when direct costs vary will be assigned a cost in the firm's decision-making process. In the terms of the model we are replacing

$$c_4(P_t - P_{t-1})^2$$

with:

$$c_4[(P_t - V_t) - (P_{t-1} - V_{t-1})]^2$$

where V_t represents the direct unit costs of production (labor, capital rental, and raw materials) in period t.[8]

[8] It may be that a more complex lag structure would be appropriate according to the speed with which increases in direct costs are recognized and transmitted into prices, but in the absence of any detailed empirical knowledge on this process and in the desire to use a single model for many industries, this structure will be retained.

Derivation of Linear Decision Rules

We can now bring the cost and revenue terms together into a single equation so that the conditions for optimization can be derived. It is perhaps best to begin by summarizing the notation. Let:

$X_t \equiv$ rate of production in period t
$P_t \equiv$ price in period t
$U_t \equiv$ level of unfilled orders (backlog) at the end of period t
$U_t^* \equiv$ desired level of unfilled orders at the end of period t
$H_t \equiv$ level of finished goods inventories at the end of period t
$H_t^* \equiv$ desired level of finished goods inventories at the end of period t
$O_t \equiv$ new orders in period t
$S_t \equiv$ shipments in period t

These variables are constrained by the following identities:

$$O_t - S_t \equiv U_t - U_{t-1}$$
$$X_t - S_t \equiv H_t - H_{t-1}$$

The demand equation relates the endogenous variable O_t to the decision variable P_t and the exogenous term Q_t:

$$O_t = Q_t - bP_t$$

We will treat the problem initially as one of certainty. Furthermore we assume that the firm discounts future profits according to a discount factor $\lambda(0<\lambda<1)$. Therefore, to maximize discounted future profits over an N-period horizon, we maximize the Lagrangian expression in equation (**A**) shown on the following page where δ_t and γ_t are Lagrangian multipliers, and λ a discount factor. After taking derivatives for any period t with respect to the decision variables and allowing t to take on

(A)

$$L = \sum_{t=1}^{N} \lambda^{t-1} \Big\{ P_t Q_t - bP_t^2 - c_{11} - c_1(U_t - c_{13} - c_{14}X_t)^2 - c_{21} - c_2(H_t - c_{23} - c_{24}S_t)^2$$
$$- V_t X_t - c_3(X_t - X_{t-1})^2 - c_4[(P_t - V_t) - (P_{t-1} - V_{t-1})]^2 \Big\}$$
$$- \delta_t(Q_t - bP_t - S_t - U_t + U_{t-1}) - \gamma_t(X_t - S_t - H_t + H_{t-1})$$

all values to N(where N is large), the system is solved according to the Z transform procedure outlined in Holt et al.[4] The end result is a set of linear decision rules for production, price, and finished goods inventory of the form:

$$(1)\quad X_t = A_{11}X_{t-1} + A_{12}P_{t-1} + A_{13}H_{t-1} + A_{14}U_{t-1} + A_{15}Q_t + A_{16}Q_{t+1} + \ldots + A_{17}V_{t-1} + A_{18}V_t + \ldots + k_1$$

$$(2)\quad P_t = A_{21}X_{t-1} + A_{22}P_{t-1} + A_{23}H_{t-1} + A_{24}U_{t-1} + A_{25}Q_t + A_{26}Q_{t+1} + \ldots + A_{27}V_{t-1} + A_{28}V_t + \ldots + k_2$$

$$(3)\quad H_t = A_{31}X_{t-1} + A_{32}P_{t-1} + A_{33}H_{t-1} + A_{34}U_{t-1} + A_{35}Q_t + A_{36}Q_{t+1} + \ldots + A_{37}V_{t-1} + A_{38}V_t + \ldots + k_3$$

where the k's are constants. There are also rules for U_t and S_t which can be derived from the first three.

Since $Q_t, Q_{t+1} \ldots$ and $V_t, V_{t+1} \ldots$ are not known quantities but random variables, the firm can do no better than maximize the expected value of profits. According to the Simon-Theil theorem, the same decision rules apply provided we substitute for Q and V their expected values, $\hat{Q}$ and $\hat{V}$. These equations are used to determine the first period's decision for X, P, and H. Each period the rule is used again, with the current period being treated as the first.

In a normative model, the firm would make a forecast of future Q and V and insert the resulting values into its decision rules. Since we do not have any information about what forecasts the firms used, for purposes of regression analysis we must guess at the specific set of forecasts on which the decisions were based. Following previous studies of this type by Childs and Holt et al., we have used $\hat{Q}_t = Q_t$, i.e., on the assumption that the firm is on average an accurate predictor of future events, the demand that actually comes about is used as a proxy for what the firm had anticipated at the time the decisions were made.[5,6] The same formulation is used for $\hat{V}$.

Aggregation

The model outlined above and the resulting equations for production, price, and finished goods inventory dealt with the decision process of the individual firm. The available data cover the industry aggregates corresponding to these variables. Problems of aggregation have been treated elsewhere (see Theil) and will not be discussed here. The main point is that the coefficients of the industry "decision

[4] The exact solution is not presented here due to lack of space.

[5] Naive and distributed lag expectations were tried in our regressions and did not do as well as the perfect forecasts.

[6] Q_t is not directly observed but can be obtained for purposes of regression analysis by the inverse relation $Q_t = O_t + bP_t$. Although b is not known either, an a priori "reasonable" value (see Section II) can be used. As a check, alternative values of b were used to generate the Q series with very little effect on the regression results. Note that for the firm Q is a true exogenous variable. The use of $(O+bP)$ as a proxy for what the econometrician cannot observe does not change the reduced-form nature of the actual decision rules.

rule" will not in general be a simple function only of the corresponding micro-coefficients. To achieve an unbiased set of restrictions would require knowledge at the individual firm level which we simply do not possess. In the absence of such knowledge the sample calculations discussed below will assume that the industry can in fact be treated as a single large firm and will assume that any restrictions on the micro-coefficients will carry over, with appropriate scaling, to the corresponding macro-coefficients.

II. *Calculation of Decision Rules for Particular Cost Structures*

This section reports on an attempt to establish some of the qualitative properties of the system which has equations (1)–(3) as its reduced form. Specifically, some estimate should be provided as to the range within which the regression coefficients are likely to lie if the observed decisions are made in a manner approximated by the decision model. Furthermore, since the A_{ij} in the reduced form are functions of all the parameters in the original structure (b, λ, and the c's), it is desirable to know how the A's will be affected by different assumptions about the values of these parameters.

> Different firms or industries to which the model might be applied have different cost parameters, and one wants to know how this fact will affect the [A's]. In addition, it is known that within a firm or industry these parameters do not usually remain constant over the period covered by a time-series sample. Thus, if the regression coefficients are very sensitive to changes in the underlying parameters, results will be poor when the regressions are estimated from empirical data even if the basic decision model is correct. [Mills, p. 139]

For a less complex structure, the coefficients of the reduced form equations could be derived analytically as functions of the parameters of that structure. With the present model, however, that approach does not appear to be feasible. It would require, in addition to matrix manipulation and determinant evaluation, the analytic solution of an eighth degree equation.

Given the infeasibility of an analytic solution, an alternative approach is attempted in the present study. The system is solved for specific values of the cost and revenue parameters. As these parameter values are changed in successive trials, the results are examined to determine the sensitivity of the A_{ij} to the specific parameters used. Hopefully, for a wide range of values for the b, λ, and c's, the A coefficients will retain at least the same signs, and to a degree, the same relative magnitudes. The extent to which the coefficients *are* sensitive to the parameters gives some indication of the results which might be expected when parameter values other than the ones used in the samples are appropriate.

Parameter Specifications

The first stage of the experiment consists of the selection of an initial set of cost and revenue parameters to be used to arrive at a specific numeric solution for the derivation of the decision rule coefficients. Including λ, the discount factor, and b, the slope of the demand curve, there are twelve parameters to be specified. The following characteristics of the data will help to set the order of magnitude for these parameters.

Price is in index form and is of the order of 100, with a standard deviation of approximately 3.8. The physical series (shipments, inventories, etc.) are reported in millions of dollars. Shipments, production, and new orders averaged approximately \$1000 million with a standard deviation of approximately \$165 million. Inventories and unfilled orders each average around \$600 million with a standard deviation of

TABLE 1—DECISION RULE COEFFICIENTS FOR A PARTICULAR COST STRUCTURE

Independent Variable	X_{t-1}	P_{t-1}	H_{t-1}	U_{t-1}	Q_t	Q_{t+1}	Q_{t+2}[a]	V_{t-1}	V_t	V_{t+1}	V_{t+2}[a]
Dependent Variable											
X_t	.260	−.028	−.345	.495	.492	.089	.019	.028	−.032	−.004	−.001
P_t	−.169	.903	−.058	.066	.156	.106	.081	−.903	1.028	−.010	−.008
H_t	.195	−.007	.354	−.001	−.001	.024	.012	.007	−.009	−.003	−.001

[a] The two infinite series were truncated at this point. This is justified by the exponentially declining weights in the expansion of $Q(Z)$ and $V(Z)$.

approximately $100 million and $80 million, respectively.[7]

With these figures in mind, the following were assumed to be the values of the cost parameters:

i) c_{11}, c_{13}, c_{21}, and c_{23} drop out entirely or affect only the intercept term and were assumed equal to zero.

ii) $b=.05$. Given the magnitude of the price and new orders data, this corresponds to a price elasticity of .5 which seems reasonable given the assumptions about the demand curve (i.e. that it specifies quantity demanded from the firm when all firms charge the same price).

iii) $\lambda=.99$. Since the decision period is one month, this corresponds to an implicit annual discount of approximately 12 percent.

iv) $c_{14}=c_{24}=.6$. Thus it is assumed that the observed long-run ratios between inventories and shipments and unfilled orders and production approximate the desired relationship.

v) $c_1=150$; $c_2=100$; $c_3=35$; $c_4=5$. It was assumed that if the actual level of any of the variables deviated from the desired (or previous period's) level by as much as a standard deviation (approximately), the cost would equal 10 percent of total revenue for that month.

The results of the initial sample calculation are presented in Table 1 above. In Tables 2-a, b, and c we attempt to provide some notion of how the A_{ij} depend on the structural parameters. In each of the eight trials a single parameter, indicated in the left-hand column, was varied from the initial set of values. In the case of λ, the value was lowered to .89. In all other cases, the parameter under consideration was doubled.

Looking ahead to the regression analysis for the production and price equations the results of the sample calculations are generally encouraging. The algebraic signs for all the coefficients remain unchanged in every trial. Furthermore, the magnitudes of many of the coefficients were relatively insensitive to changes in the parameters indicating that time-series analysis may yield meaningful results, even though the cost and revenue structure of the firms involved have undoubtedly undergone some degree of change over the time period studied. For the inventory equation, however, the sample calculations produce ambiguous results, with only the signs of X_{t-1} and H_{t-1} remaining the same in every trial.[8]

It is interesting to note that the production and inventory rules are generally unaffected by different specifications regarding the cost of changing price, c_4, and the

[7] There are some small differences between the two industries studied for all of these figures, but not enough to warrant separate treatment.

[8] This is only partially shown in Table 2. Many other parameter specifications were tried but are not reported here.

slope of the demand curve, b. Similarly, the rule for price is relatively insensitive to c_{14} and c_{24}. As evidence of the interdependence among decisions, however, we note the strong impact of c_3, the cost of changing the production rate, on the price decision rule.

In terms of individual coefficients it is worth noting that the coefficients of U_{t-1} and Q_t are virtually identical in the pro-

TABLE 2-a: PRODUCTION DECISION RULE X_t: CHANGE FROM INITIAL SET UNDER ALTERNATIVE COST STRUCTURES

Value of Parameter which is changed	X_{t-1}	P_{t-1}	H_{t-1}	U_{t-1}	Q_t	Q_{t+1}	Q_{t+2}	V_{t-1}	V_t	V_{t+1}	V_{t+2}
$\lambda=.89$	+.003	−.001	0	+.005	+.005	−.004	−.001	+.001	+.001	0	0
$c_1=300$	−.053	−.001	−.026	+.053	+.053	−.021	−.006	+.001	−.001	0	0
$c_2=200$	−.021	−.001	−.028	+.046	+.046	+.012	+.002	+.001	0	−.001	0
$c_3=70$	+.096	+.004	+.078	−.132	−.132	−.006	+.003	−.004	+.007	−.001	−.001
$c_4=10$	−.003	−.001	−.001	+.001	+.002	+.002	+.002	+.001	+.001	−.001	0
$c_{14}=1.2$	−.111	+.006	+.033	−.021	−.020	−.031	−.005	−.006	+.006	+.002	0
$c_{24}=1.2$	−.011	−.005	+.036	+.090	+.090	+.015	−.002	+.005	−.004	−.001	0
$b=.10$	+.015	−.023	+.007	−.008	−.010	−.005	−.003	+.023	−.025	−.004	0

TABLE 2-b: PRICE DECISION RULE P_t: CHANGE FROM INITIAL SET UNDER ALTERNATIVE COST STRUCTURES

Value of Parameter which is changed	X_{t-1}	P_{t-1}	H_{t-1}	U_{t-1}	Q_t	Q_{t+1}	Q_{t+2}	V_{t-1}	V_t	V_{t+1}	V_{t+2}
$\lambda=.89$	−.013	+.032	+.001	+.002	+.005	−.004	−.008	−.032	−.017	0	+.001
$c_1=300$	−.004	0	−.003	+.006	+.007	+.001	0	0	0	0	0
$c_2=200$	−.009	−.001	0	+.002	+.002	+.001	0	+.001	0	0	0
$c_3=70$	−.135	−.005	−.039	+.040	+.039	+.014	+.004	+.005	−.004	−.002	−.001
$c_4=10$	+.079	+.030	+.031	−.035	−.079	−.052	−.038	−.030	−.005	+.005	+.004
$c_{14}=1.2$	−.050	−.001	+.024	−.024	+.027	−.005	0	+.001	+.002	0	0
$c_{24}=1.2$	+.035	+.001	+.001	+.002	+.002	−.005	−.004	−.001	−.003	0	0
$b=.10$	−.135	−.052	−.051	+.060	+.045	+.011	−.006	+.052	−.007	−.012	−.007

TABLE 2-c: INVENTORY DECISION RULE H_t: CHANGE FROM INITIAL SET UNDER ALTERNATIVE COST STRUCTURES

Value of Parameter which is changed	X_{t-1}	P_{t-1}	H_{t-1}	U_{t-1}	Q_t	Q_{t+1}	Q_{t+2}	V_{t-1}	V_t	V_{t+1}	V_{t+2}
$\lambda=.89$	−.001	0	−.001	+.005*	+.004*	+.001	−.001	0	0	+.001	0
$c_1=300$	+.006	+.001	0	−.002	−.002	+.007	+.001	−.001	+.001	0	0
$c_2=200$	−.058	−.002	−.057	+.098*	+.097*	+.006	−.002	+.002	−.001	+.001	0
$c_3=70$	+.079	+.004	+.070	−.115	−.116	−.019	−.001	−.004	+.006	+.001	−.001
$c_4=10$	−.001	0	0	0	0	0	+.001	0	0	0	0
$c_{14}=1.2$	−.054	+.001	−.065	+.116*	+.116*	+.021	+.008	−.001	0	+.002	0
$c_{24}=1.2$	−.009	−.009	+.072	+.207*	+.205*	+.012	−.007	+.009	−.009	0	0
$b=.10$	+.004	−.004	+.002	−.001	−.002	−.002	−.001	+.004	−.003	−.002	−.001

* Indicates a sign change.

duction equation and the inventory equation. (This was true on all trials.) Thus for purposes of these two decisions, an order on hand at the beginning of the period is treated the same as an order anticipated during the period. This result is a consequence of the particular cost structure assumed since Q_t and U_{t-1} enter the cost function symmetrically through the constraint:

$$Q_t - bP_t - S_t = U_t - U_{t-1}$$

Childs (pp. 36–37) obtained the same result in a similar model (without a price equation) and suggests that it demonstrates the possible over-simplification in the U_t^* equation, whereby the desired level of unfilled orders was assumed to depend solely on current production. He suggests that one alternative would be to make U_t^* a function of O_t as well as X_t which would serve to break the rigid equality between the two coefficients, although we would not expect the signs to change. More importantly, he recommends that in the regressions, the coefficients not be constrained to have the same sign (i.e., estimated as a single variable), since this would almost certainly involve some degree of specification error.

Alternatively we might think of this result as arising from our neglect of the production-to-order, production-to-stock distinction. If, as we mentioned earlier, the desired level of unfilled orders is related not to total production but only to that portion of the total which is production-to-order, then the equality between the coefficients would be broken, with U_{t-1} receiving a smaller weight, especially in the inventory equation. However, U_{t-1} would still appear in the rule for finished goods inventories so long as the cost of changes in the rate of production is a function of total production and independent of the mix between order and stock.

Even assuming the plausibility of the original cost structure, this result serves to demonstrate one of the more questionable implications of the certainty equivalent-quadratic criterion function approach. U_{t-1} is a known quantity, given to the decision maker before he must make his decisions on X_t and H_t. Ignoring the perfect forecast, $\hat{Q}_t$ is not a known quantity but the expected value of the decision maker's subjective probability function over all possible values of Q: yet for purposes of decisions on the rate of production and the level of finished goods inventory, they are treated identically. Of course, this is a straightforward result of the certainty equivalence theorem, and, perhaps more suggestively, a consequence of assuming a quadratic criterion function. The point is that this approach in some ways seems to beg the question of risk and uncertainty, and, in so doing, may lead us to overlook what some regard as the essential characteristic of unfilled orders—its ability to serve as a buffer between the past and an uncertain future.

We also note that in the price equation, the coefficients of H_{t-1} and U_{t-1} are opposite in sign and almost equal in absolute value. Hence, our model suggests that for purposes of the price decision, it may be possible to consider unfilled orders as "negative" inventory. However, as before, this result derives from the simple aggregation of order and stock production with a single price for both. In the regression analyses, we would expect the coefficients of H_{t-1} and U_{t-1} to partly reflect the relative weights of order and stock goods in the aggregate price variable, and therefore not display the same coefficient.

Finally we note that P_{t-1} and V_{t-1} have the same absolute coefficients in all equations and the relevant variable for the model is therefore $(P_{t-1} - V_{t-1})$. Because the P series is in index form, however, we do not constrain the variables to have the same coefficients; nor should we expect

them to turn out equal if they are not constrained.

III. *Regression Analysis*

The Data

The data consist of monthly observations for the period March 1953—August 1966 on two Standard Industrial Classification (*SIC*) manufacturing groups—Lumber and Wood Products (*SIC* 24), and Paper and Allied Products (*SIC* 26).

The primary sources for the data are manufacturer's shipments, inventories and unfilled orders series compiled by the Industry Division of the Bureau of Census, *U.S.* Department of Commerce,[9,10] and the Wholesale Price Index and average hourly earnings series compiled by the Bureau of Labor Statistics, *U.S.* Department of Labor.

As is generally known, the Wholesale Price Index is commodity-oriented, i.e., the index is the weighted sum of the prices for a group of similar commodities. On the other hand, the Commerce Department series are industry-oriented, i.e., the shipments series, for example, will include all the shipments by a particular industry despite the fact that the specific commodities which originate in a single industry may be quite diverse. Therefore the coverage of a typical Commerce series is likely to be considerably different from that of the most nearly related price series.

For the industries examined in this study, the coverage of the two types of index is reasonably close. Indeed this was the main reason for selecting these particular industries. Hopefully, therefore, the errors introduced by the coverage problems will be minimized, although one's confidence in the specific numerical results is somewhat weakened. It should be mentioned that generally one can expect a closer correspondence of the two types of data at lower levels of aggregation. Unfortunately, although many of the data are available at the 3-digit level, the series for inventories by stage of fabrication is currently available only for 2-digit aggregates.

A further problem arises in attempting to obtain estimates for the direct costs of production. These are to be used in testing the amended specification of the cost of changing price in which penalties are associated with price changes that are correlated to changes in direct costs. Although labor costs can be represented by the series on average hourly earnings, there is considerable difficulty in arriving at a similar index for raw material costs. The problem arises when the output series at the 2-digit level includes as final product its own basic raw material input. For example, the principle raw material input for Lumber and Wood Products (according to *BLS* weights) is lumber. Yet lumber also gets added in as final product, and is, in fact, the largest item in that aggregate. Hence the largest component in the price index for the final product is identical to the largest component in the index of raw material prices and we can expect a rather substantial spurious correlation between the two indices. Therefore, labor costs were the only element of direct costs used in the regression analysis. Also, due to the high degree of collinearity in the series, only values for periods $t-1$ and t were used.

Discussion of Results

Regressions were performed using simple least squares on both seasonally adjusted

[9] Not all the data are published by Commerce. I am grateful to David Belsley for making available to me the series which he obtained under the stipulation that only the results derived from them and not the series themselves be published. For a detailed description of the Commerce data, see Belsley's doctoral dissertation.

[10] Since the theory is in terms of physical quantities, the shipments, inventories and unfilled orders variables which are reported by Census in value terms were deflated by the price variable for the corresponding month. This procedure introduces a slight degree of error if businesses do not value finished goods inventory at finished goods prices.

TABLE 3—REGRESSION COEFFICIENTS
(t-values in parentheses)

	α	X_{t-1}	P_{t-1}	H_{t-1}	U_{t-1}	Q_t	Q_{t+1}	Q_{t+2}	V_{t-1}[a]	V_t	$\bar{R}^2$*	$D.W.$[b]
LUMBER												
X_t	−3.620	.213	−.004	−.106	.157	.568	.074	.008	−.006	.015	.904	2.03
		(3.99)	(0.34)	(1.59)	(3.18)	(9.47)	(1.24)	(0.16)	(0.47)	(1.04)		
P_t	5.584	−.261	.914	−.318	.063	.130	.172	.366	−.012	.002	.947	1.42
		(2.73)	(39.26)	(2.65)	(0.72)	(1.21)	(1.62)	(4.13)	(0.47)	(0.07)		
H_t	−0.890	−.008	.012	.914	−.017	−.074	.042	.017	.014	−.011	.902	1.62
		(0.26)	(1.63)	(23.98)	(0.59)	(2.16)	(1.25)	(0.59)	(1.74)	(1.34)		
PAPER												
X_t	−1.490	.023	−.043	−.300	.265	.854	−.060	−.056	.107	−.085	.979	1.98
		(0.75)	(4.34)	(3.16)	(6.40)	(26.53)	(1.96)	(2.03)	(5.56)	(4.10)		
P_t	−0.040	−.019	.970	−.554	.123	.086	−.026	.049	−.031	.046	.981	1.80
		(0.31)	(49.20)	(2.90)	(1.48)	(1.32)	(0.42)	(0.89)	(0.80)	(1.12)		
H_t	−0.250	.024	.004	1.003	−.056	.005	.041	.006	.031	−.035	.982	1.59
		(1.56)	(0.80)	(20.67)	(2.66)	(0.32)	(2.61)	(0.39)	(3.11)	(3.31)		

[a] Average hourly wages in cents.
[b] These are presented without inference as to their significance, given the presence of a lagged dependent variable in the equations. In addition, David Grether has pointed out that the relevant serial correlation may be other than first order for monthly data, e.g., $E(\epsilon_t, \epsilon_{t+n}) \neq 0$.
* 150 degrees of freedom.

and nonadjusted data. The latter is reported below although there was very little difference between the two. The results, which are presented in Table 3, are generally consistent with the expectations generated by the sample calculations. Indeed, none of the 20 coefficients which were predicted unambiguously by the sample calculations showed the wrong sign for both industries. Overall, out of 40 such coefficients, 34 were predicted correctly. Even more revealing, out of 18 coefficients which came out with coefficients more than twice their standard error, only 1 was incorrect. The same regressions without an intercept term were performed on first differences. Although the $\bar{R}^2$'s dropped considerably, the matching of coefficients with expectations was not much changed.

When the results are broken down by equation they become somewhat more meaningful. The production equation performed quite well, showing the predicted sign for 14 of 18 coefficients and for 9 of the 10 coefficients which turned out significant. The coefficient of X_{t-1} is somewhat smaller than had been predicted, especially for the paper industry. This indicates that production may be somewhat more flexible than had been assumed. If c_2 were made smaller, the effect would be to lower the coefficient of X_{t-1} and shift more of the weight to U_{t-1} and Q_t. It is rather surprising that in the regressions on first differences, the coefficient of X_{t-1} comes out significantly negative for both industries. This result could not be reproduced in the simple calculations by any reasonable set of cost parameters and one is tempted to attribute it to some statistical problems (e.g., estimating first differences without an intercept).

The price equation is also strong with 17 of 18 signs predicted correctly, and of the

6 coefficients which turned out significant, all had the correct sign. The price equation is rather heavily dominated by the lagged price term, and this may reflect the tendency of the *BLS* index to pick up list prices which do not necessarily reflect the terms at which transactions actually take place. To the extent to which this is a quantitatively significant phenomenon we would not expect other variables to show up significantly since in the extreme $P_t = P_{t-1}$ for long periods, even though actual transactions prices are changing in response to market conditions.

The coefficient of H_{t-1} in the price equation is considerably larger than predicted, and this result holds up in the first difference regressions. No set of cost parameters could be found which would significantly improve the predictions of the simple calculations for this coefficient without throwing off some of the others; for example, increasing b makes the coefficient of H_{t-1} larger (in absolute value) but moves the coefficient of X_{t-1} in the wrong direction. This suggests that there may be some specification error involved here.

There is little that can be said about the inventory equation since the predictions for most signs were ambiguous. If we judge by the results of the initial sample calculation, 14 of 18 coefficients were predicted correctly, including all 7 which came out significant. In one sense the results can be considered favorable since the model predicted that most coefficients in the inventory equation would have values very close to zero and this was borne out. The extremely high coefficient of H_{t-1} was unexpected, but since this result does not hold up for the first difference equations, there is little value in trying to come up with an economic interpretation. Since the Durbin-Watson statistics suggest serial correlation of the residuals in the inventory equation, the value of H_{t-1} is subject to bias, and this might explain the result.

TABLE 4—THE IMPACT OF A ONE UNIT INCREASE IN DEMAND

	Lumber	Paper	Predicted (TABLE 1)
Increase in production	.568	.854	.492
Rise in price	.006	.004	.008
Drawing down of inventories	.074	−.005	.001
Buildup of unfilled orders	.352	.147	.499

Impact of Changes in Demand

It is interesting to ask how the system reacts to changes in demand; specifically, how the impact is spread over the different variables the firm controls—production, price, inventory, unfilled orders, and shipments. The answer is dependent on the coefficient of Q_t in the various equations together with the identity constraints involving those variables. The information is summarized in Table 4 above. Each number indicates what portion of a sudden and temporary one unit increase in demand, Q, would be borne by the variable indicated in the left-hand column.[16] (The amount absorbed by a price increase is found by multiplying the coefficient of Q_t in the price equation by .05, the assumed slope of the demand curve.)

It is perhaps surprising that price change plays such a small role in absorbing increases in demand although that is precisely what is predicted by the model.[17] To some extent this is attributable to the previously mentioned failure of the *BLS* index to reflect all price changes, but it also serves to point out a possible deficiency in

[16] If the increase in demand is foreseen three months in advance, the adjustment process is only slightly different.

[17] Table 4 understates the total contribution of price change even where the rise in demand is unexpected since price continues above its long-run equilibrium level for several periods. The full effect of price increases is to absorb 11, 4, and 7 percent of the demand increase in lumber, paper, and the sample experiment, respectively.

the model. The assumption that all cost parameters remain constant throughout the cycle is extremely convenient mathematically, but almost certainly misrepresents the facts to some degree. For example, as capacity is approached it must be true that c_3 (the cost of increasing production) rises sharply, and it is at this point that price increases are most likely to play an important role. If the model could be amended to include factors such as capacity utilization, the true role of price changes would probably show up more emphatically.

IV. *Summary and Conclusions*

In this paper we have attempted to develop a model which explains decisions on production, price, and finished goods inventory for manufacturing firms. The keynote of the analysis was that decisions on these variables should be treated as simultaneous and interdependent, emphasizing that price should not be considered as given, but rather as one of the variables to be determined within the model.

To what extent can we consider the model a "success?" It is difficult to attach a score to the overall contribution of the analysis but some observations are possible. First, we have demonstrated that it is technically possible to treat the decision variables as being determined simultaneously, and that given certain assumptions with respect to the demand curve, there is no reason why the price decision cannot be included. Second, the model has the desirable property that it is capable of generating empirically testable (and therefore refutable) hypotheses under most conditions. This may explain why as yet there has been relatively little enthusiasm for many of the nonmaximization theories which, although perhaps appealing to our intuitive feelings about the workings of the business world, have not, for the most part, yielded substantive implications with regard to observable phenomena. Finally, the model seems to hold up reasonably well when applied to selected 2-digit industry groups, despite obvious problems with the data. In particular we noted the imperfect matching of the *BLS* price series to the *SIC*-based production series, and the possibility that even the "correct" price list might not always reflect the actual terms of a transaction.

The final point to be made regards the direction of future research in this area. Certainly there is still much work to be done within the framework of the current model. Some different specifications might be tried, for example making H_t^* depend on new orders as well as shipments and similarly for U_t^*. Alternative lag structures could be tried for the direct cost variables and the treatment of expectations about demand might be improved. We might even want to get behind the production variable by actually including a production function in the model. This might enable us to derive decision rules for the size of the work force, amount of raw materials, and inventories at various stages of fabrication. Finally, we might expand the model to include some of the less easily quantified variables such as advertising and other aspects of nonprice competition.

In many ways, however, the most critical need is for better data. The individual problems have been discussed above and it is not necessary to repeat them here. However the point must be stressed that it is only with data which are accurate, consistent, and appropriate to the level of aggregation of the model that any theory in this area can be given a fair and rigorous test.

REFERENCES

D. A. Belsley, "Industry Production Behavior: An Economic Analysis," unpublished doctoral dissertation, M.I.T. 1965.

G. L. Childs, *Unfilled Orders and Inventories: a Structural Analysis*, Amsterdam 1967.

P. Darling, and M. Lovell, "Factors Influencing Investment in Inventories," in Dusenberry et al., eds., *The Brookings Quarterly Econometric Model of the U.S.*, Chicago 1965.

O. Eckstein, and G. Fromm, "The Price Equation," *Amer. Econ. Rev.*, Dec. 1968, *43*, 1159–83.

W. Fellner, *Competition Among the Few*, New York 1960.

C. C. Holt, F. Modigliani, J. F. Muth, and H. A. Simon, *Planning Production, Inventories, and Work Force*, Englewood Cliffs 1960.

E. S. Mills, *Price, Output and Inventory Policy*, New York 1962.

B. R. Moss, "Industry and Sector Price Indices," *Mon. Lab. Rev.*, U.S. Department of Labor, Aug. 1965.

H. A. Simon, "Dynamic Programming under Uncertainty with a Quadratic Criterion Function," *Econometrica*, Jan. 1965, *24*, 74–81.

H. Theil, *Linear Aggregation of Economic Relations*, Amsterdam 1954.

———, "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica*, Apr. 1957, *25*, 346–49.

Returns from Investment in Human Capital

By THOMAS JOHNSON*

A number of economists have attacked the problem of personal income distribution by considering the nonproperty earnings of an individual to be the returns to the varied mixture of human resources which the individual possesses. These resources may be broadly classified as raw labor and human capital. Several recent studies have analyzed investments in human capital by examining the internal rate of return to schooling.[1] These estimates compute the internal rate of return from discrete investments with returns received at discrete points in time.

Besides the general theoretical difficulties associated with using the internal rate of return as an investment decision tool, there are some empirical difficulties with the method as it has been employed to analyze the relationships between different investments in human capital. With discrete investments and returns, the internal rate of return must be computed by iterative schemes and is not amenable to tests of hypotheses. Also, observed incomes are not predicted well when the computed internal rate is used to compute the expected (constant) return per period from observable investments in schooling.

The objective of this paper is to develop a model which will predict lifetime earnings. It is formulated such that parameters can be estimated simultaneously and hypotheses tested (at least approximately) for several types of investment in human capital, as a function of race and region. The model is developed by representing an individual's investment in human capital as a continuous process, rather than a series of discrete investments. It includes a base earning capacity which is considered to be the return on the initial endowment of human resources at the time the individual makes the decision to invest in human capital. Investments in human capital are then measured in terms of the fraction of earnings capacity which is foregone at each moment in time. In addition, depreciation of human resources and autonomous growth in earnings with time are included in the model in general functional form. The basic model is an integral equation which is solved for the age profile of earnings capacity. Of course if depreciation of human resources and autonomous growth are of the same functional form, then their effects will be confounded if we use only a single cross-section of earnings data.

It is assumed that the rate of return is constant for all investments in human capital, that the rates of depreciation and growth are constant, and that the fraction of earning capacity invested in human capital is 1.0 while the individual is in school and declines linearly from an estimated value at the end of schooling to zero at the 65th birthday (assumed retirement). Even with these simplifying assumptions,

* The author is assistant professor in the department of economics, Southern Methodist University. He wishes to express his appreciation for helpful comments and criticisms from T. D. Wallace, R. M. Fearn, L. A. Ihnen, and participants in the Econometrics and Human Resources Workshops at North Carolina State University. All remaining errors are, of course, his alone.

The research for this paper was supported, in part, by National Science Foundation Grant GS-2391 while the author was a research associate in the department of economics, North Carolina State University.

[1] See for example, G. Becker (1964, 1967), Hanoch, W. L. Hansen, B. R. Chiswick, Becker and Chiswick, and J. Mincer.

the model is non-linear and there is no apparent linearization transformation. Thus, non-linear regression techniques are used.

The data which are employed for estimation of the parameters are the earnings profiles which Giora Hanoch (1965, 1967) estimated from the 1960 one-in-one-thousand sample survey. These are age-earnings profiles in cross-section for males by region (North and South) and race (white and nonwhite). With these data the model is assumed to represent the earnings of the typical individual in each race/region group. The empirical results should be compared with those obtained by Hanoch from the same data, and should be considered indicative of the capabilities of the model rather than the final word for policy applications.

The coefficients of determination between the "observed" and calculated earnings are mostly on the order of 0.95, with from seven to thirteen degrees of freedom although values as low as 0.87 and as high as 0.99 are obtained. In every case investigated, the rate of return computed in the model of this study is higher than the internal rate of return which Hanoch computes from the same data. The estimated "net" rates of return (the estimated rate of return less the estimated depreciation rate) more nearly correspond to the estimates of the internal rate of return. However, most of the net rates are also larger than the internal rates and the use of a single cross-section of data biases the net rate downward by the rate of autonomous growth in earnings.

Of course the model has limitations: The work-leisure choice is not explicitly analyzed. Rather it is hidden in the schedule of supply of human capital services which underlies the concept of the rate of return as a price for these services. The retirement choice is not analyzed but retirement is assumed to occur on the 65th birthday. Property and other nonearnings income are excluded from the model although they obviously affect the investment decisions. Analysis of these choices would require a reformulation of the basic structural form.

I. *The Model*

General Functional Form

The *discrete* models of income generation (see Gary Becker 1967, Becker and B. R. Chiswick, and Chiswick) represent the earning capacity of a typical individual as the sum of the base earnings which he would receive without investing in human capital plus the returns on his previous investments in human capital. In *integral* form this process may be written as

$$(1) \quad E(t, \beta) = B(t, \beta) + \int_0^t R(y, \beta)k(y, \beta)E(y, \beta)dy$$

where

- t = the "age" of the individual in years measured from the date at which the investment decision is made;
- β = the vector of characteristics of the group of which the individual is typical. These characteristics are independent of time (this does not include occupation);
- $E(t, \beta)$ = the individual's earning capacity at age t;
- $B(t, \beta)$ = the base earnings capacity which the individual with characteristics β would have at age t;
- $R(t, \beta)$ = the return the individual having β characteristics will receive per unit time from a unit investment in human capital made at age t; and
- $k(t, \beta)$ = the fraction of earning capacity at age t which is invested in human capital at t for the individuals having characteristics β.

Equation (1) is a Volterra's Integral equation of the second kind.[3]

If the stock of human resources depreciates at the rate $\delta(\mathrm{t}, \beta)$, depending on the time at which the depreciation occurs but not on the time at which the investment was made, and if this depreciation is the only way in which the base earnings capacity varies during the life of the individual, the base earning is

$$B(\mathrm{t}, \beta) = B(\beta) \exp\left[- \int_0^t \delta(u, \beta) du \right]$$

where $B(\beta)$ is the earnings capacity at $\mathrm{t}=0$. The derivation of this result from the depreciation of the initial stock of human resources is given in T. Johnson. Similarly, an investment in human capital, $k(\mathrm{y}, \beta)E(\mathrm{y}, \beta)$ made at time y has, at $\mathrm{t}>\mathrm{y}$ depreciated to

$$k(\mathrm{y}, \beta)E(\mathrm{y}, \beta) \exp\left[- \int_y^t \delta(u, \beta) du \right]$$

and at time t the earnings capacity per unit time from this investment is

$$R(\mathrm{y}, \beta)k(\mathrm{y}, \beta)E(\mathrm{y}, \beta) \cdot \exp\left[- \int_y^t \delta(u, \beta) du \right]$$

Now at time t the earning capacity is the sum of the "base" earning capacity and the earnings capacity of the previous investments in human capital. Thus,

$$E(\mathrm{t}, \beta) = B(\beta) \exp\left[- \int_0^t \delta(u, \beta) du \right] + \int_0^t R(\mathrm{y}, \beta)k(\mathrm{y}, \beta)E(\mathrm{y}, \beta) \cdot \exp\left[- \int_y^t \delta(u, \beta) du \right] d\mathrm{y}$$

In my dissertation it is shown that the solution of this equation for $E(\mathrm{t}, \beta)$ is given by

$$(2) \quad E(\mathrm{t}, \beta) = B(\beta) \cdot \exp\left\{ \int_0^t [R(u, \beta)k(u, \beta) - \delta(u, \beta)] du \right\}$$

Up to now the vector of characteristics, β, has been carefully included in the argument of variables. As more specific functional forms are used, this notation will be dropped but the idea that these characteristics enter into the function will be retained.

With the form of the model given in equation (2) the connection with more familiar results is more easily seen for a special case. Thus, consider the case in which the rate of return and the rate of depreciation are both independent of time, i.e., $R(\mathrm{t})=R$ and $\delta(\mathrm{t})=\delta$. With this assumption that the returns per unit time are equal for all investments, earning capacity can be given by the relation developed by Y. Ben-Porath, $E(\mathrm{t})=RH(\mathrm{t})$, which also leads to

$$(3) \quad E(\mathrm{t}) = B \exp\left\{ \int_0^t [Rk(u) - \delta] du \right\}$$

Since the integrand in equation (3) may be written as $R[k(u)-\delta/R]$ it is seen that, with R and δ constant, $k(\mathrm{t})-\delta/R$ may be thought of as the net investment fraction. Alternatively, it is shown in Johnson that if δ is a constant, $R(\mathrm{t})-\delta$ may be thought of as the net rate of return to the investment $k(\mathrm{t})E(\mathrm{t})$. If we also set $\delta=0$ and $k(\mathrm{t}, \beta)=1$, then

$$E(\mathrm{t}, \beta) = B(\beta) \exp (R\mathrm{t})$$

which says that if all earnings are continuously reinvested at a constant rate, earning capacity grows at a continuous, constant rate, R.

The effect of cross-sectional observations, when lifetime earnings streams are growing autonomously, can now be put in terms of the model. Only the effects and

[3] See F. B. Hildebrand.

not the causes of such autonomous growth are investigated. Let $E(\theta, \beta, t)$ be the earning capacity at time t of the cohort of age θ (in 1960, say). If there is growth in $E(\theta, \beta, t)$ at rate $g(\theta)$, then

$$\frac{dE(\theta, t)}{d\theta} = -g(\theta)E(\theta, t) \tag{4}$$

The sign is generally understood to be negative since positive growth in earnings capacity over time means that for larger θ (older persons in 1960) the earning capacity, E, at age t is lower than for smaller θ at the same age. If $E(\theta, t)$ and $g(\theta)$ are integrable functions of θ, the solution of equation (4) is

$$E(\theta, t) = E(0, t) \exp\left[-\int_0^\theta g(w)dw\right]$$

With "cohort age" defined as the t age in 1960, taking equation (2) to be $E(0, t)$, the cross-sectional earnings in 1960 are

$$E(t, t) = \exp\left[-\int_0^t g(w)dw\right]B \cdot \exp\left\{\int_0^t [R(u)k(u) - \delta(u)]du\right\}$$

This last result above shows that if the growth rate and the depreciation rates are such that $\delta(t)+g(t)$ is of the same functional form as $\delta(t)$ and $g(t)$ (e.g., polynomials of equal degrees), then growth biases only the estimate of depreciation with cross-sectional observations. Otherwise any secular variation of the growth rate would affect the estimates of parameters other than δ.

Cross-sectional observations of earnings capacity in the year $1960+v$ (where t is the age in $1960+v$ so that $\theta = t - v$) are given by

$$E(t - v, t) = B \exp\left[-\int_0^{t-v} g(w)dw\right] \cdot \exp\left\{\int_0^t [R(u)k(u) - \delta(u)]du\right\} \tag{5}$$

Equations (2) and (5) present the model in a very general form.[3] The development of the model with these general functional forms for $R(t)$, $k(t)$, $\delta(t)$, and $g(t)$ gives it great potential for representing the earnings stream of the typical individual of a group. However, general functional forms cannot be implemented empirically. Thus, a second stage of development of the model is required to put the model in a parametric form which can be implemented empirically.

The Model with Simplifying Assumptions

From equations (2) and (5), it is seen that the time dependent rate of return, $R(t)$, and the time dependent function for investment in human capital, $k(t)$, appear only in the form $R(t)k(t)$. Thus, in the general functional form the effects of changes in rates of return and changes in investment fractions (on-the-job training (OJT) after schooling is completed) are not distinguishable. In order to distinguish the effects of the rate of return and investments in OJT, it is assumed that the rate of return is the same for all investments in human capital by the representative individual, i.e., $R(t) = R$, and for $k(t)$ a functional form is assumed to distinguish between schooling and OJT investments.

The assumption of a single rate of return is rationalized on the basis that essentially marginal decisions are analyzed. It is assumed that the principal decision concerning schooling is to extend the time spent in a school of a fixed quality. Although commitments for extension of schooling are made in (sometimes lengthy) blocks of

[3] Thus far the model has been developed to represent the earnings per unit time which the individual receives at a point in time. If instead it is desirable to model the average income of a group of individuals over an extensive period of time, the expression for Y can be integrated and used to estimate the model parameters from data on the average earnings over broad age intervals if the population distribution is known.

time, it may be argued that shorter intervals of schooling are not considered when the decision is made to invest in the schooling. Furthermore, if the average rate of return does not change with length of the schooling period after a given age, then during that period the marginal rate is equal to the average rate. The marginal decision concerning OJT is to vary intensity (i.e., the size of $k(t)$) at each time, and the rate of return is the marginal price. With perfect knowledge, the income maximizing individual would equate the marginal rates of return. However, this is only approximate since no correction is made for finite life.

Next, it is necessary to assume a tractable form for the fraction, $k(t)$, of earning capacity invested in human capital. Becker (1967, p. 75) finds that approximately 26 percent of the private costs of college are tuition and other direct costs. Becker then assumes that college students earn 25 percent of their annual opportunity wage through summer work. On this basis it is assumed that while an individual is in school he invests his entire earning capacity in schooling, i.e., $k(t)=1$. This assumption is also made by Hanoch (1965, 1967) on the same justification.

After schooling is completed, it is assumed that the gross investment fraction is a linear function of time from the end of schooling, $t=S$, until retirement, $t=T$. In addition, Ben-Porath argues that gross investment expenditures, $k(t)E(t)$, are expected to be greater than zero for $t<T$. Thus, the following functional form is assumed for the fraction, $k(t)$, of earning capacity invested in human capital:

$$(6)\quad k(t) = \begin{cases} 1.0 & t \leq S \\ A - \dfrac{A}{(T-S)}(t-S) & t > S \end{cases}$$

where A is the fraction of earnings capacity invested in OJT immediately after schooling is completed,[4] and it is assumed that $0<A<1$.

For the lack of any compelling theoretical reason for a complicated functional form for $\delta(t)$ and $g(t)$, it is assumed that both $\delta(t)$ and $g(t)$ are constant. From the above analysis it is seen that from cross-sectional data $\delta+g$ will be estimated. For this reason, and to simplify notation, it is assumed that $g(t)=0$.

It seems reasonable to define net (observed) earnings to be earnings remaining after gross investment expenditures, or

$$(7)\qquad Y(t) = (1-k(t))E(t)^{5}$$

With this definition and the simplifying assumptions, we have

$$Y(t) = 0, \quad \text{when} \quad t < S$$

$$(8)\ Y(t) = B\left[1 - A + \frac{A}{(T-S)}(t-S)\right] \cdot \exp\left[(R-\delta)S + (RA-\delta)(t-S) - \frac{RA(t-S)^2}{2(T-S)}\right] \quad \text{when} \quad t > S$$

Alternatively, one might define observed earnings to be earnings before expenditures for maintenance of the stock of human capital and seek an estimate of these expenditures. In this study the first, less demanding assumption has been used. It is further assumed that for $t>S$ all investments in human capital are accounted for by foregone earnings so that for $t>S$, $Y(t)$ is the observed earnings.

With these specific assumptions con-

[4] A more general linear form of $k(t)$, $t>S$ was tried. This form was A_1+A_2 (t-S), but problems of identification of the model were encountered. An attempt to estimate T in equation (8) would encounter the same limitation as the estimation of A_2. For a discussion of the problems of identification in this model, see my dissertation.

[5] Of course, it is understood that earnings are assumed to contain no physical capital component.

cerning the parametric form of the model, several quantities may be derived as closed form functions of these parameters. The total investment in schooling after the decision is made to continue ($t=0$) is given by $\int_0^S E(t)dt$. The total (gross) investment in OJT is given by $\int_S^T k(t)E(t)dt$. With discounting at the rate ρ the present value at $t=0$ of the base earnings stream if no additional investment is made is given by $\int_0^T Be^{-(\delta+\rho)t}dt$. Similarly, the present value of the net earnings stream $Y(t)$, when S additional years of schooling are taken is given by $\int_S^T Y(t)e^{-\rho t}dt$. Also, the amount of depreciation of human capital during schooling and OJT may be calculated as $\delta/R \int_0^S E(t)dt$ and $\delta/R \int_S^T E(t)dt$, respectively. These integrals are evaluated in the appendix to my dissertation.

The Shape of Observed Earnings

It is interesting to examine the earnings profiles implied by equation (8). The maximum of $E(t)$ is found by setting $\partial E(t)/\partial t=0$ which, assuming $B\neq 0$, requires that $R(t)k(t)-\delta(t)=0$. That is, net investment must be zero for $E(t)$ to attain a maximum at $t<T$. With the simplifying assumptions this also implies that $E(t)$ is a maximum at $t_{max}=S+(A-\delta/R)\cdot((T-S)/A)$.

Now computing $dY(t)/dt$ from equation (7), when $\partial E(t)/\partial t=0$,

$$\frac{dY(t)}{dt} = -E(t)\frac{dk(t)}{dt}$$

It may be seen that $Y(t)$ is still increasing when $E(t)$ is a maximum, because $k(t)$ is expected to be a decreasing function of time.

These characters of the earnings profiles, $E(t)$ and $Y(t)$, agree with the forms anticipated by Ben-Porath. It will be shown below that the parameters of the model can be estimated and that observed cross-sectional earnings profiles are well approximated.

Interpretations of the Parameters in the Model

Each of the parameters in the model has an economically meaningful interpretation. The parameter B is the return per unit *time* on all of the human resources which the individual possesses at the time he makes his decision. Thus, B depends on the "raw labor," "ability," and "human capital" possessed at that time by the investor. If, at some age, the typical individuals in two subgroups (say white and nonwhite) of a cohort group have equal raw labor, and no human capital, the observed differences in the B's between them would give a market measure of the difference in ability between the two subgroups. Of course, such a measure of ability must be interpreted with caution because at any age at which individuals have a positive earning capacity much has been invested in them. Prenatal and infant medical care is one such form of early investment in individuals as is the informal education imparted by parents.[6]

From the basic formulation, equation (1), it is seen that R is the "price" or "rent" which the individual receives per period for the services of his stock of human capital. The stock of human capital possessed is determined by the values of the entire set of parameters. There is no requirement in the model as it stands to require that R be an interest rate which equates the present value of two income streams. Thus, R is not an internal rate of return but a measure of the rent actually received per unit of investment.

The parameter A (in equation (8)) is

[6] The empirical results of this study do not support this type of measurement of ability or cultural disadvantage. The discussion is presented as a theoretical interpretation of the parameter B. Hopefully, future work with better data will be able to utilize this parameter more fully.

the fraction of earnings capacity invested in *OJT* immediately upon leaving school. With the restriction that the fraction of earnings capacity invested in *OJT* declines linearly to zero at a fixed retirement age, the parameter A determines the history of investment in *OJT*.

We will now turn our attention to estimating the parameters of the simplified model, using data derived from the 1960 Census one-in-one-thousand sample.

II. *Estimation of the Model Parameters*

Data

Estimates of the parameters B, A, R, and δ are derived from earnings profiles derived by Hanoch (1965, 1967) from the 1960 Census one-in-one-thousand sample. These data may be expected to yield somewhat different results than the unadjusted data used in most previous studies. The numerical results should be taken with the several reservations occasioned by the prior adjustments to the basic census data.

Hanoch has used regression analysis on characteristics of males fourteen years and older in this sample to obtain corrected earnings versus age profiles for several levels of schooling for whites and nonwhites and the North and South. Hanoch used seventy-eight variables (sixty characteristic, or dummy, variables and eighteen continuous) to "correct" observed earnings for the characteristics of the individual. A group was defined by a qualitative characteristic (e.g., nonwhite), or by an interval of a continuous variable (e.g., age 25–34). Some variables, such as schooling age, and weeks worked, were represented in both a continuous and an alternate dummy variable form.

Hanoch smoothed the earnings-age profiles by taking ten-year moving averages.[7] With the recognition that these are not observations of the earnings of a typical individual in a race-region-schooling group they are used in the estimation of the model *as if* they were such observations. Since this is a single cross-section, cohorts are not considered and the values are taken to correspond to net earnings, $Y(t)$.

Hanoch (1965, p. 62) is careful to emphasize that ". . . the functions estimated here adjust the schooling differential for other factors within the particular group, but they do not adjust for any differences among the groups." He also argues that "Hence, the comparisons are valid if restricted to a given Race/Region. They would not be adequate for comparisons of similar individuals *between* the races or the regions, because the covariates were not held constant between these groups."

Despite this limitation, and the fact that the earnings at only selected ages are tabulated, Hanoch's results seem to be useful for estimating the model developed above. The important thing to remember is that race and regional comparisons include effects of differences in the covariates used by Hanoch. It would certainly be interesting to obtain more pure regional and racial comparisons after correcting for intergroups differences in the covariates. However, it is believed that the results obtained with Hanoch's tabulated data are sig-

[7] With these manipulations of the earnings-age observations, when the resulting values are used as observations in the logarithmically transformed model developed above, it is assumed that the errors of these observations are distributed as *log*-normal. Hanoch does not make the logarithmic transformation for the following reasons: he was interested in absolute earnings; he found that the logarithmic transformation overcorrected for the skewness of the distribution of earnings; and he was not particularly concerned with bias in the estimates of the variances. More refined analysis of the distribution of estimates with the model in general is a potential area for future research. Another possibility is to consider the properties of a fully coordinated stepwise process where the observed earnings (or *log*-earnings) are first regressed on the non-age variables (linearly) and then the model used with nonlinear regression on the residual to account for age.

nificant without further refinements of correction for covariates.

The Statistical Model

Income distributions viewed cross-sectionally have generally been found to have a skewed distribution approximately *log*-normal where the distribution has been tested. After removing systematic components, it is assumed that the residual cross-sectional error is similarly skewed and approximately distributed as a *log*-normal. In census data of median incomes of a group of given age there is but one observation at each age. Thus, with this type of data the operative assumption will be that the residual error is log-normally distributed with the same distribution at each age in the cross-section. On the basis of this tradition, for estimation, equation (8) is transformed by taking natural logarithms.

With the assumption of a multiplicative error term which is distributed *log*-normally and independently for each observation of $Y(\mathrm{t})$, for $\mathrm{t}>\mathrm{S}$ with $X_i=\mathrm{t}_i-\mathrm{S}$, and applying the logarithmic transformation $Y^*(\mathrm{t})=log_e Y(\mathrm{t})$ yields the form estimated:[8]

$$(9)\quad \begin{aligned} Y_i^*(\mathrm{t}) = log_e B[1 - A + AX_i/(\mathrm{T}-\mathrm{S})] \\ + (R-\delta)\mathrm{S} + (RA-\delta)X_i \\ - RAX_i^2/(\mathrm{T}-\mathrm{S}) + \epsilon_i \end{aligned}$$

where $\epsilon \sim N(0, I\sigma^2)$

[8] Since we are interested in estimating B, A, R, and δ, these parameters are retained instead of a different set obtained by further reparametrization. The conventional approach is taken in assuming the error after transformation is distributed with zero mean and constant variance. More refined considerations of bias in the estimates of the original parameters are not investigated. D. M. Heien has developed an unbiased estimate of the constant term for a model which is linear in the logarithms. An extension to analysis of models which are still non-linear after the logarithmic transformation is a potential area for future research. It is recognized that this assumption about the errors is not the most refined. However, the development of a clearly superior error assumption in this non-linear situation has not been attempted.

The model has been developed in terms of the earnings profile of an individual. As usual, however, the hypotheses which are of interest concern the difference between groups with special characteristics. Thus, some aggregation across individuals is not only acceptable but desirable. By taking the model as developed to represent the earnings of the typical individual in the group, aggregation is immediate when the observations are representative of the group.

In the form of equation (9), the model is still non-linear (there is no apparent transformation to make the model linear in the parameters) and thus standard linear regression techniques cannot be used to estimate the parameters or test hypotheses. N. R. Draper and H. Smith (ch. 10) give a good introduction to the techniques of non-linear estimation. The non-linear regression method which was used to estimate the parameters of equation (9) was suggested by H. O. Hartley. A discussion of the computer program used is given by A. C. Nelson et al. and Johnson. The elements of the variance-covariance matrix of the least-squares estimates are approximated by the elements of the matrix $V^{-1}\sigma^2$ where $2V=[\mathrm{S}_{ij}]$ is the matrix of second partial derivatives of the sum of squared deviations. See George Box and G. Coutie and E. M. L. Beale.

Because of compulsory school attendance laws, child labor laws, and the decreasing number of individuals with very few years of schooling, the movement from the 0–4 year level to the 5–7 year level is not analyzed as the result of an economic decision. Thus the first decision analyzed is the decision to move from the 5–7 year level to the 8 year level of schooling. The ages which Hanoch used for the completion of schooling levels were also used here.

Retirement, corresponding to T, is

assumed to be on the sixty-fifth birthday. Thus, for a given lower level and a given upper level of schooling, the value T−S is 65 minus the birthday age at which earnings start. Similarly, the X_t variable in equation (9) is the birthday age minus the birthday age at the first year out of school. Investigation of the identification of equation (9) according to the criterion of Fisher, shows that T cannot be estimated together with B, A, R, and δ.

Estimating the four parameters B, A, R, and δ of equation (9) with these data yields results which indicate a relation between B and the other parameters which prevents precise estimates of B simultaneously with other parameters.[9] On the basis of these results, B is not estimated in the regressions. Instead, independent estimates of B are entered as constants in the equation for each race/region/schooling group. A description of the manner in which estimates were obtained for B is given in the Appendix. The estimates of B are, therefore, not reliable.

III. *Results*

The results of the regressions to estimate A, R, and δ are tabulated in Table 1. The approximate standard deviation of each estimate is given in parentheses immediately following the value estimated. These approximations of the variances are obtained with the second-order approximation outlined in footnote 8. The coefficients of determination, R^2, cannot be calculated using the usual formulas since in the non-linear regression there is no requirement that the mean of the estimates equal the mean of the observations. Instead, I used the basic concept that the coefficient of determination is the square of the correlation between the observed and predicted values of the dependent variable.

Among the results tabulated in Table 1, a large number of comparisons can be made and hypotheses formulated. In comparing results between races and regions it is seen that in most cases both the rate of return, R, and the fraction invested in OJT are higher for whites in the South than for whites in the North, and are higher for nonwhites in the South than for whites in the South.

In keeping with the objective to develop a model in which hypotheses can be tested, the comparisons where lower rates of return are estimated for nonwhites in the South than for whites in the South are chosen to test the hypothesis $H_0;R_{w/S} = R_{nw/S}$.[10] The comparisons between the groups designated in Table 2 are chosen for testing of hypotheses.

In each case, the hypothesis tested is that the rates of return, R, are equal between Group 1 and Group 2 as designated in Table 2. To test the hypothesis, the parameters for both groups are estimated simultaneously with the model entered in the program such that the rates of return are required to be equal. Thus, in the restricted regressions, five parameters are estimated (two A's two δ's, and one R). The F-values are calculated from the following formula

$$F = \frac{(SSE_{\text{RESTRICTED}} - (SSE_1 + SSE_2))/k}{(SSE_1 + SSE_2)/(DF_1 + DF_2)} \tag{10}$$

where:

$SSE_{\text{RESTRICTED}}$ = sum of squares of residuals in the regression when the restrictions are applied;

[9] This is the same type of identification problem which is the result of multicollinearity in a linear model. The magnitudes of the approximate variances of the estimates of B are so large that refined tests are not required to indicate that it is better not to estimate B in the regressions.

[10] These tests appear to be "results guided." However, they are guided by the direction of the deviations and not by the magnitude of the deviations.

k = number of parameters restricted in value in the combined regression as compared to the separate regressions;

SSE_i = sum of squares of residuals for the unrestricted regression for group i ($i=1, 2$). Taken from Table 1; and

DF_i = degrees of freedom in the unrestricted regression for group i ($i=1, 2$). Taken from Table 1.

Under the assumption that the F of equation (10) is distributed approximately as the standard central F, the computed values are compared with tabulated critical values at 5 percent and 1 percent levels of probability with k and DF_1+DF_2 degrees of freedom. It should be reemphasized that the probability levels are only approximate since the distribution of the F in equation (10) has not been established.

The last comparison in Table 2 shows that the significance of differences between schooling levels may be tested as well as between races and regions.[11] This particular test shows that nonwhite graduate students in the South earn a rate of return on their investment which is significantly different (lower) than the rate earned by nonwhite college graduates.[12]

Table 3 presents some additional quantities which are implied by the values of the parameter estimates given in Table 1. In Table 3, $R-\delta$ is the net rate of return. Comparisons with the internal rate of returns calculated by Hanoch (1967) from the same data show that in only four cases is the estimate of $R-\delta$ slightly less than Hanoch's estimate of the internal rate of return. The total observed income from the end of schooling until retirement is given as $\int_S^T Y(t)dt$, and very consistent values are obtained for persons completing a given grade level for the different decision points analyzed (e.g., 16 years completed with the decision to continue either at 12 or 13–15 years). Investments in schooling and OJT are calculated from the closed form formulas developed in Johnson. Net investments have the amount of depreciation during the particular investment phase subtracted from the gross amount invested. The calculated investment amounts display less internal consistency than total earnings because they are less directly related to the observed income streams from which the parameter estimates were obtained. The measures of gross investment in schooling and OJT are much higher than Mincer estimated for 1958. However, Mincer could not identify depreciation and his measures are more comparable to net measures.

In examining the values in Table 3 it should be remembered that the estimate of the depreciation rate, δ, is biased by the use of cross-sectional data. The estimate of δ affects the estimates of the net investment amounts and the age at which net investment is zero as well as the net rate of return, $R-\delta$.

Conclusions

The results presented in Table 1 indicate that cross-sectional average earnings streams for large groups are well represented by the model with a constant rate of return to investments in human capital, a constant rate of depreciation (plus

[11] Although no such tests are made here, it would be easy to test for differences in the model parameters between race/region and schooling/decision groups.

[12] One explanation for this difference might be occupation differences with graduate students going into less well paid occupations such as teaching and the ministry.

Table 1—Estimates of Model Parameters with "Independent" Estimates of B[a]

Race/Region	Schooling Levels Lower	Upper	B	A[b]	R	δ	SSE	DF	R^2
	5–7	8	1,750	.630	.297	.0934	.30607	13	.956
				(.028)	(.0120)	(.0055)			
	8	9–11	1,900	.480	.321	.0750	.15683	12	.958
				(.026)	(.0120)	(.0043)			
	8	12	1,900	.430	.275	.0599	.05949	11	.973
				(.018)	(.0064)	(.0029)			
	9–11	12	2,510	.384	.311	.0590	.06001	11	.973
				(.019)	(.0094)	(.0030)			
White/North	12	13–15	3,920	.486	.240	.0648	.01643	9	.988
				(.013)	(.0043)	(.0018)			
	12	16	3,920	.479	.217	.0618	.01279	8	.992
				(.014)	(.0034)	(.0018)			
	13–15	16	7,160	.517	.197	.0630	.01321	8	.989
				(.014)	(.0042)	(.0019)			
	12	17+	3,920	.490	.187	.052	.07550	7	.991
				(.016)	(.0027)	(.0017)			
	16	17+	10,770	.561	.157	.054	.06586	7	.992
				(.015)	(.0035)	(.0015)			
	5–7	8	1,450	.626	.284	.0875	.36165	13	.994
				(.021)	(.0130)	(.0060)			
	8	9–11	1,580	.486	.337	.0826	.11302	12	.975
				(.021)	(.0100)	(.0037)			
	8	12	1,580	.450	.288	.0650	.05746	11	.979
				(.017)	(.0062)	(.0029)			
	9–11	12	2,260	.424	.309	.0648	.05785	11	.974
				(.018)	(.0086)	(.0029)			
White/South	12	13–15	3,420	.522	.277	.0820	.09233	9	.973
				(.028)	(.0099)	(.0043)			
	12	16	3,420	.569	.268	.0944	.15135	8	.965
				(.038)	(.0110)	(.0061)			
	13–15	16	6,760	.608	.245	.0952	.15508	8	.964
				(.040)	(.0130)	(.0061)			
	12	17+	3,420	.508	.219	.0658	.01762	7	.986
				(.022)	(.0039)	(.0026)			
	16	17+	11,550	.623	.162	.0688	.01470	7	.990
				(.021)	(.0048)	(.0023)			
	5–7	8	1,000	.631	.287	.082	.63616	13	.916
				(.041)	(.0170)	(.0250)			
	8	9–12	920	.355	.379	.061	.40964	12	.876
				(.042)	(.0250)	(.0071)			
	8	12	920	.346	.348	.064	.12059	11	.953
				(.023)	(.0098)	(.0042)			
	9–11	12	1,170	.268	.455	.063	.12176	11	.953
				(.022)	(.0180)	(.0042)			
Non-white/North	12	13–15	2,340	.448	.240	.006	.02325	9	.986
				(.017)	(.0054)	(.0022)			
	12	16	2,340	.436	.261	.093	.19376	8	.969
				(.051)	(.0130)	(.0072)			
	13–15	16	3,910	.439	.258	.093	.19361	8	.969
				(.056)	(.0180)	(.0072)			
	12	17+	2,340	.634	.316	.133	.05525	7	.994
				(.027)	(.0630)	(.0043)			
	16	17+	5,760	.580	.352	.132	.05619	7	.994
				(.035)	(.0130)	(.0044)			

TABLE 1—(*Continued*)

Race/Region	Schooling Levels Lower	Upper	B	A[b]	R	δ	SSE	DF	R^2
	5–7	8	1,000	.636	.271	.090	.33217	13	.946
				(.030)	(.0130)	(.0058)			
	8	9–11	1,060	.526	.350	.106	.40299	12	.947
				(.038)	(.0180)	(.0069)			
	8	12	1,060	.567	.342	.121	.15975	11	.983
				(.023)	(.0093)	(.0047)			
Non-white/South	9–11	12	1,600	.559	.348	.121	.15906	11	.984
				(.025)	(.0120)	(.0046)			
	12	13–15	2,640	.533	.125	.046	.03612	9	.954
				(.023)	(.0064)	(.0028)			
	12	16	2,640	.617	.268	.133	.08519	8	.993
				(.027)	(.0078)	(.0046)			
	13–15	16	3,230	.568	.297	.132	.09027	8	.993
				(.031)	(.0110)	(.0047)			
	12	17+	2,640	.658	.137	.058	.01252	7	.986
				(.018)	(.0034)	(.0021)			
	16	17+	5,660	.740	.110	.060	.01281	7	.981
				(.015)	(.0042)	(.0020)			

[a] These estimates are based on Hanoch's (1965) estimates of earning streams for males in 1959, computed from regressions within age groups, smoothed by moving averages of ten years. Earnings of students are assumed equal to direct private costs. These estimates represent private costs and returns.

[b] Numbers in parentheses are second order approximations to the standard deviations of the parameters. Note that theoretically the parameter A should be the same for a given scholastic group. The different estimates of A shown in the table for different base schooling levels are another indication of variability in the estimates. In most cases, the estimated values of the parameters imply that gross investment continues to increase after schooling is complete. This is consistent with the case analyzed by Ben-Porath (p. 361) where capital accumulation reduces the cost of producing human capital.

TABLE 2—TESTS FOR EQUALITY BETWEEN SELECTED RATES OF RETURN FOR RACE/REGION/SCHOOLING GROUPS

Group 1	Group 2	Null Hypothesis	DF	F
Whites/South Lower schooling level = 5–7 Upper schooling level = 8	Nonwhites/South Lower schooling level = 5–7 Upper schooling level = 8	$R_1=R_2$	1/26	0.527 4.22[a] (7.72)
Whites/South Lower schooling level = 12 Upper schooling level = 13–15	Nonwhites/South Lower schooling level = 12 Upper schooling level = 13–15	$R_1=R_2$	1/18	178 4.45 (8.40)
Whites/South Lower schooling level = 12 Upper schooling level = 17+	Nonwhites/South Lower schooling level = 12 Upper schooling level = 17+	$R_1=R_2$	1/14	208 4.60 (8.86)
Whites/South Lower schooling level = 16 Upper schooling level = 17+	Nonwhites/South Lower schooling level = 16 Upper schooling level = 17+	$R_1=R_2$	1/14	66.7 4.60 (8.86)
Whites/South Lower schooling level = 12 Upper schooling level = 16	Nonwhites/South Lower schooling level = 12 Upper schooling level = 17+	$R_1=R_2$	1/15	222 4.54 (8.68)

[a] Values in parentheses are the 5 percent and 1 percent points for the standard F distribution.

TABLE 3—QUANTITIES DERIVED FROM PARAMETER ESTIMATES

Race/Region	Lower Schooling Level	Upper Schooling Level	$R-\delta$	$\int_S^T Y(t)dt$	Gross Schooling Investment	Net Schooling Investment	Gross OJT Investment	Net OJT Investment	Birthday at which Net Investment=0	Net Positive OJT Investment[a]
White/ North	5–7	8	.2032	$199,200	$ 4,318	$ 2,958	$ 91,562	$− 28	40.5	$18,903
	8	9–11	.2456	220,025	4,907	3,760	68,335	902	42.1	15,316
	8	12	.2146	245,433	12,036	9,408	67,664	− 690	42.1	14,704
	9–11	12	.2521	245,399	6,528	5,291	57,951	453	42.8	13,235
	12	13–15	.1752	278,323	15,471	11,293	93,916	− 6,598	41.7	17,170
	12	16	.1554	341,856	38,863	27,800	115,698	−14,548	41.8	18,216
	13–15	16	.1346	341,935	26,463	18,047	130,252	−19,921	41.0	18,679
	12	17+	.1349	380,626	56,445	40,738	129,256	−12,627	44.0	23,033
	16	17+	.1027	380,486	23,912	15,682	160,077	−25,980	42.3	23,300
White/ South	5–7	8	.1964	$162,795	$ 3,552	$ 2,458	$ 73,232	$ 524	40.9	$15,663
	8	9–11	.2542	180,462	4,118	3,108	58,271	− 303	41.3	12,195
	8	12	.2229	221,344	10,201	7,890	64,520	− 253	42.3	14,113
	9–11	12	.2438	221,322	5,825	4,603	59,201	334	42.7	13,304
	12	13–15	.2148	259,226	13,937	9,803	98,603	− 7,520	41.1	16,842
	12	16	.1733	311,244	36,090	23,364	141,187	−18,347	40.9	19,480
	13–15	16	.1502	311,527	25,619	15,678	158,538	−23,867	40.1	20,069
	12	17+	.1461	338,222	51,927	35,807	125,412	−18,511	42.4	18,339
	16	17+	.0932	337,394	25,394	14,611	175,842	−42,337	39.8	17,856
Non-white/ North	5–7	8	.2048	$141,147	$ 2,472	$ 1,766	$ 60,205	$ 2,739	42.8	$14,550
	8	9–11	.3187	138,410	2,573	2,162	28,359	1,671	43.8	7,322
	8	12	.2841	155,129	6,850	5,585	33,442	− 1,392	41.0	6,587
	9–11	12	.3926	155,092	3,555	3,065	24,286	− 402	41.9	5,158
	12	13–15	.1747	146,957	9,227	6,708	46,396	− 6,406	39.4	6,793
	12	16	.1674	142,428	24,185	15,536	50,344	−18,591	33.0	1,808
	13–15	16	.1652	142,378	15,182	9,705	50,859	−18,860	32.9	1,801
	12	17+	.1828	258,916	42,442	24,558	149,577	−22,553	40.4	16,448
	16	17+	.2202	258,811	14,473	9,057	127,700	−16,956	41.1	15,410
Non-white/ South	5–7	8	.1807	$ 91,128	$ 2,409	$ 1,608	$ 44,022	$− 934	39.4	$ 8,545
	8	9–11	.2437	91,562	2,731	1,905	36,382	− 2,365	38.0	5,826
	8	12	.2204	101,437	6,804	4,386	48,144	− 5,004	36.8	6,271
	9–11	12	.2263	101,458	4,046	2,634	47,032	− 4,794	36.9	6,186
	12	13–15	.0784	97,513	8,929	5,622	39,477	−11,266	35.8	3,694
	12	16	.1346	105,415	24,373	12,247	66,493	−19,034	33.5	2,839
	13–15	16	.1650	105,283	12,539	6,964	57,156	−15,064	34.5	2,990
	12	17+	.0797	127,890	29,548	17,157	69,323	−13,382	41.4	8,961
	16	17+	.0508	127,833	11,915	5,484	88,285	−28,366	38.0	6,665

[a] Net positive OJT investment is the integral of net OJT investment from the end of schooling to the time at which net investment is zero.

autonomous growth), the single parameter representation of the fraction of earnings capacity invested in OJT, and the other assumptions (e.g., retirement at the sixty-fifth birthday, and investment of the total earnings capacity in human capital while the person is in school). Also, the model and estimation procedure provide for testing hypotheses concerning relationships between groups of the values of these parameters. Thus the objective of model development has been met.

APPENDIX

Estimates of the Base Earning Capacity

Reasonable estimates of B the base earn-

ing capacity for males having obtained each level of schooling for each race/region group are obtained in the following manner. All four parameters, B, A, R, and δ were estimated simultaneously for the groups who decide at age fourteen, after completing five to seven years of schooling, to continue schooling for two years and for whom the estimates of B are intuitively reasonable. In these regressions the estimate of A was fairly steady at a value of 0.6. It was then assumed that 0.6 of earning capacity is invested in *OJT* at age fourteen by the groups who quit school and start work at that age so that 0.4 of earning capacity is observed earnings. The base earning capacity, B, of this group is then approximately (observed earnings at age fourteen) divided by 0.4. These are assumed to be the earning capacities at age fourteen for all males in each race/region group who acquire at least five to seven years of schooling. These values of B are then used as constants in estimating A, R, and δ for the groups who continue schooling through the eighth grade and then quit.

Estimates of the base earning capacity of individuals who decide to continue schooling on the completion of eight years are obtained by dividing the observed earnings for the first year of work of those who quit school after eight years by one minus the estimated value of A. With the base earnings capacity obtained in the above manner, the parameters, A, R, and δ, are then estimated with the regression program for those persons deciding to complete nine to eleven or twelve years of schooling. After these estimates of A are obtained, estimates of the base earnings capacity of individuals who decide to continue schooling after completion of nine to eleven or twelve years are calculated in the same manner as for earlier decisions.

While this method of estimating the B's does not obtain truly independent estimates of the B's, it has compensating advantages. First, the B estimates obtained are intuitively more reasonable than those obtained when all four parameters are estimated in the regressions. Second, going back to the data for observed earnings at each decision age reduces the number of ways in which a progressive divergence from the observed earnings can occur. In addition, the use of the 0.6 estimate for the first A resulted in little change of the estimated parameters from the values which were obtained when all four parameters, B, A, R, and δ, were estimated by the non-linear regression. Thus the subjective confidence in these estimates of B is high.

REFERENCES

E. M. L. Beale, "Confidence Regions in Non-linear Estimation," *J. Royal Statist. Soc.*, 1960, No. 1, *B-22*, 41–76.

G. Becker, *Human Capital*, New York 1964.

———, *Human Capital and the Personal Distribution of Income*, Ann Arbor 1967.

——— and B. R. Chiswick, "Education and the Distribution of Earnings," *Amer. Econ. Rev.*, May 1966, *56*, 350–69.

Y. Ben-Porath, "The Production of Human Capital and the Life Cycle of Earnings," *J. Polit. Econ.*, Aug. 1967, *75*, 352–65.

G. E. P. Box and G. A. Coutie, "Application of Digital Computers in the Exploration of Functional Relationships," *Proc. Institution of Electrical Engineers*, Apr. 1956, *B103*, 100–07.

B. R. Chiswick, "Human Capital and the Distribution of Personal Income," unpublished doctoral dissertation, Columbia Univ. 1967.

N. R. Draper and H. Smith, *Applied Regression Analysis*, New York 1966.

F. M. Fisher, *The Identification Problem in Econometrics*, New York 1966.

G. Hanoch, "Personal Earnings and Investment in Schooling," unpublished doctoral dissertation, Univ. Chicago 1965.

———, "An Economic Analysis of Earnings and Schooling," *J. Hum. Resources*, summer 1967, *2*, 310–29.

W. L. Hansen, "Total and Private Rates of Returns to Investment in Schooling," *J. Polit. Econ.*, Apr. 1963, *71*, 128–40.

H. O. Hartley, "The Modified Gauss-Newton Method for the Fitting of Non-linear Regression Functions by Least Squares," *Technometrics*, May 1961, *3*, 269–80.

D. M. Heien, "A Note on Log-Linear Regression," *J. Amer. Statist. Ass.*, Sept. 1968, *63*, 1034–38.

F. B. Hildebrand, *Methods of Applied Mathematics*, 2d ed., Englewood Cliffs 1965.

T. Johnson, "Returns from Investment in Schooling and On-the-Job Training," unpublished doctoral dissertation, N.C. State Univ., Raleigh 1969.

J. Mincer, "On-the-Job Training Costs, Returns and Some Implications," *J. Polit. Econ.*, Oct. 1962, *70*, 50–79.

A. C. Nelson, Jr., C. A. Krohn, J. R. Batts, W. S. Thompson and T. Johnson, *Evaluation of Computer Programs for System Performance Effectiveness*, Research Triangle Inst., Research Triangle Park, N.C. 1967.

F. Welch, "Labor-Market Discrimination: An Interpretation of Income Differences in the Rural South," *J. Polit. Econ.*, June 1967, *75*, 225–40.

U.S. Bureau of the Census, *Census of Population: 1960*. Final Report PC (2)-5B. Washington 1963.

Measurement of Portfolio Performance Under Uncertainty

By IRWIN FRIEND AND MARSHALL BLUME*

Harry Markowitz's pioneering work in portfolio theory and James Tobin's subsequent extension forced a complete reevaluation of the pricing of capital assets under uncertainty. A by-product of this reevaluation was a theory of equilibrium in the capital markets, which was independently discovered by William Sharpe (1964), John Lintner (1965a), and Jack Treynor (1965). The theory, which in this paper will be called the "market-line" theory, led to several different, although related, one-parameter measures of the investment performance of an asset or a portfolio.

The entire rationale of one-parameter measures of investment performance is to replace two-parameter measures of performance—rate of return and risk—with a single measure which uses market data to combine the two different dimensions of performance into a single measure which adjusts for differences in risk. A single risk-adjusted measure of performance is not only simpler than a combination of risk and return measures, but permits, at least theoretically, a definitive comparison of performance of investments with different returns and risks. It is not surprising, therefore, that one-parameter measures of performance are receiving more and more attention, but it is surprising that there has been virtually no statistical analysis of the extent to which the risk-adjusted rates of return successfully abstract from risk.

The usefulness of these one-parameter measures depends, of course, upon the validity of the asumptions underlying the market-line theory. Section I of this paper briefly reviews this theory, including the assumptions on which it is based, and discusses the different one-parameter measures of performance that have been derived from the theory. Section II examines the adequacy of the one-parameter measures of performance by measuring empirically the relationship between these measures and the risk from which they are supposed to abstract. It then attempts to explain the apparent discrepancies between the market-line theory and the empirical findings in terms of specific deficiencies in the underlying assumptions.

The importance of these one-parameter measures of performance, and the associated theory of equilibrium in the capital markets, lies not only in their usefulness for analyzing investment management and market efficiency—areas of investigation to which they have already been applied—but also in their relevance and potential utility for cost of capital problems.[1] It is, therefore, essential that a careful investigation be made of the validity of these

* The authors are professor of economics and finance and associate professor of finance, respectively at the University of Pennsylvania. They wish to thank Deno Papageorge, now of the Bank of New York, for his statistical assistance and the Oppenheimer Foundation for its financial support.

[1] Since it is typically easier to estimate the cost of capital for the market as a whole than for an individual corporation, the reasonableness of the estimated cost of capital for a corporation can be tested against the equilibrium relation between the rate of return on an individual security and the rate of return for the market. See equation (1) in Section I where $E(\tilde{R}_i)$ is a measure of the "cost" of equity on an individual security and $E(\tilde{R}_m)$ the cost of equity for stocks as a whole.

measures—a goal towards which this paper is a small first step.

I. *One-Parameter Performance Measures and Their Rationale*

The key assumptions underlying the market-line theory are 1) Every investor is a one-period expected utility maximizer and exhibits diminishing marginal utility of terminal wealth.[2] 2) All investors have the same one-period time horizon. 3) Every investor feels that he can evaluate a portfolio solely in terms of the mean and dispersion or variance of one-period returns.[3] 4) It is assumed that there are no transaction or information costs, that the borrowing and lending rates are equal and the same for all individuals, and that investors will only select portfolios with optimal combinations of risk and return. 5) All investors hold identical or homogeneous expectations about the distributions of future returns. 6) The capital market is in equilibrium.[4]

Under these assumptions, Sharpe and Lintner[5] have shown that the expected return for asset or portfolio *i*, $E(\tilde{R}_i)$, is related to the expected return on the market portfolio,[6] $E(\tilde{R}_m)$, by the equation:

$$E(\tilde{R}_i) - R_f = \beta_i[E(\tilde{R}_m) - R_f] \tag{1}$$

where R_f is the risk-free rate for borrowing or lending and β_i is defined as the $\mathrm{Cov}(\tilde{R}_i,\tilde{R}_m)$ divided by $\mathrm{Var}(\tilde{R}_m)$. The variable β_i is a measure of systematic or nondiversifiable risk. The tilde superscript indicates a random variable.

If the equilibrium relationship contained in equation (1) held for all assets on an *ex ante* basis, there would be no opportunity for abnormal profit: All assets would be correctly priced. Only if some of the above assumptions did not strictly hold for all investors and for all securities could there be incorrectly priced securities. In reality, these assumptions are not likely to hold completely, but equation (1) may, nonetheless, be an adequate approximation of reality for most securities. Yet, to explicitly recognize that not all securities are in equilibrium, equation (1) can be rewritten as

$$E(\tilde{R}_i) - R_f = \eta_i + \beta_i[E(\tilde{R}_m) - R_f] \tag{2}$$

where η_i is a measure of disequilibrium. If η_i equals zero, the portfolio or asset is in equilibrium. If η_i is greater than zero, the expected return is larger than one would anticipate on the basis of the equilibrium relationship. In the market terminology, this would represent an undervalued security. If η_i is less than zero, the security is overvalued.

[2] Eugene Fama (1970) has shown under perfect capital markets and very general assumptions about consumer behavior that an individual who actually faces a multiperiod decision would act as if he followed assumption 1).

[3] The adequacy of this assumption hinges upon how closely the individual's subjective distributions of returns can be described by two-parameters. The development in this paper assumes that the variance of these distributions is defined, which is the traditional assumption. Michael Jensen (1969) shows that his one-parameter measure can be developed with the variance undefined.

The empirical evidence of this approximation to a two-parameter distribution is conflicting: Fred Arditti finds for individual securities on the New York Stock Exchange (*N.Y.S.E.*) that at least three parameters are necessary to describe these distributions, but the empirical evidence of Blume suggests that the distribution of returns for well-diversified portfolios of *N.Y.S.E.* stocks can be very accurately approximated by two-parameter stable distributions. This latter finding is, of course, more relevant to an investor who would typically hold a well-diversified portfolio.

[4] There is a substantial body of literature pertaining to the question of equilibrium. For a recent bibliography, the reader is referred to Jensen (1969).

[5] Fama (1968) has recently shown that the two developments are for all essential purposes identical.

[6] The market portfolio is defined theoretically as the portfolio consisting of all wealth whose return is uncertain. Wealth here is construed quite broadly. Yet, in every known empirical use of the model, the return on the market portfolio has been measured by some index of common stocks listed on the *N.Y.S.E.* This practice will be followed in this paper in the absence of feasible alternatives.

The one-parameter measures of investment performance of Sharpe (1966), Treynor (unpublished), and Jensen (1968) follow from equation (2). The simplest measure is Jensen's, which is merely η_i. Since this measure is in the same units as $E(R_i)$ or R_f, which are rates of returns, it can be interpreted as the rate of return above and beyond that justified by the equilibrium relationship (1), an easily interpretable measure of investment performance.

Treynor's measure follows from (2) by dividing both sides of the equation by β_i to obtain

$$(3)\qquad \frac{E(\tilde{R}_i) - R_f}{\beta_i} = \frac{\eta_i}{\beta_i} + [E(\tilde{R}_m) - R_f]$$

The ratio on the left of the above equation is Treynor's measure. If η_i equals zero, Treynor's measure is equal to $[E(\tilde{R}_m) - R_f]$, a term which is independent of the level of the systematic risk β_i. Further insight into Treynor's measure follows from the rewriting of (3) as

$$(4)\qquad \frac{\eta_i}{\beta_i} = \frac{E(\tilde{R}_i) - R_f}{\beta_i} - [E(\tilde{R}_m) - R_f]$$

Since $E(R_m) - R_f$ is a constant, Treynor's measure is merely a translation of Jensen's measure divided by the systematic risk β_i. The dimension of Treynor's measure is therefore return per unit of systematic risk.

The derivation of Sharpe's measure from equation (2) proceeds by replacing β_i by its definition to obtain

$$(5)\qquad E(\tilde{R}_i) - R_f = \eta_i + \frac{\mathrm{Cov}(\tilde{R}_i, \tilde{R}_m)}{\sigma^2(\tilde{R}_m)} [E(\tilde{R}_m) - R_f]$$

or noting that $\mathrm{Cov}(\tilde{R}_i, \tilde{R}_m) = \rho(\tilde{R}_i, \tilde{R}_m) \cdot \sigma(\tilde{R}_i)\sigma(\tilde{R}_m)$,

$$(6)\qquad E(\tilde{R}_i) - R_f = \eta_i + \frac{\rho(\tilde{R}_i, \tilde{R}_m)\sigma(R_i)}{\sigma(\tilde{R}_m)} [E(\tilde{R}_m) - R_f]$$

If portfolios i are only taken to be efficient portfolios, Sharpe (1964) has shown that the market-line theory implies that $\rho(\tilde{R}_i, \tilde{R}_m)$ equals one. Noting this property and dividing by $\sigma(\tilde{R}_i)$, one has

$$(7)\qquad \frac{E(\tilde{R}_i) - R_f}{\sigma(\tilde{R}_i)} = \frac{\eta_i}{\sigma(\tilde{R}_i)} + \frac{E(\tilde{R}_m) - R_f}{\sigma(\tilde{R}_m)}$$

The ratio on the left of (7) is Sharpe's measure. Similarly to Treynor's measure, Sharpe's is a translation of Jensen's measure divided by the standard deviation of return. An important difference between Sharpe's measure and the previous two is that whereas Jensen's and Treynor's measures can be used for any portfolio as well as individual securities. Sharpe's measure should only be applied to portfolios which are purported to be efficient.

These three one-parameter measures of investment performance were developed in terms of *ex ante* values, which at first glance might present a difficult problem in using the measures to evaluate performance. Jensen, however, has shown that at least for his measure it is possible to obtain unbiased estimates of η_i, providing β_i and R_f are constant over time.[7] If these two parameters are constant, (2) can be rewritten in *ex post* or historical data as

$$(8)\qquad R_{it} - R_f = \eta_i + \beta_i[R_{mt} - R_f] + \epsilon_{it}$$

where R_{it} is the return for portfolio or asset i in period t, R_{mt} is the market return in period t, and ϵ_{it} is a disturbance term whose expectation is zero and which is independent of R_{mt}.[8] The constant η_i

[7] Jensen (1968) examines the biases in the estimates of η_i for a specific type of non-stationarity in the risk measure.

[8] Because the market return includes asset i, the disturbances cannot be independent of R_{mt}. For a

can be estimated by standard regression techniques.

Although (8) follows from (2) providing the risk free rate is constant, Jensen (1968) in some of his empirical work allows the risk free rate to change over time, so that R_f in (8) would be subscripted by t.

Assuming again that the risk free rate is constant, one can use (8) to derive a consistent estimate of Treynor's measure using *ex post* data. If T observations are used in estimating the parameters and if $\hat{\eta}_i$, $\hat{\beta}_i$, and $\hat{e}_{it}$ are the corresponding least-squares estimates, equation (8) can be summed over T and averaged to obtain

$$\text{(9)} \qquad \overline{R_i} - R_f = \hat{\eta}_i + \hat{\beta}_i[\overline{R_m} - R_f]$$

where the bar indicates an average. Upon dividing through by $\hat{\beta}_i$, one has on the left a consistent estimate of Treynor's index.

Sharpe's measure also follows from (9) if, in addition, one assumes that $\sigma(R_i)$ is constant over time. The mathematical development is virtually the same as that which was used in deriving (7) from (2).

In the empirical part of this paper, the three performance measures will be estimated using monthly data. In estimating Sharpe's and Treynor's measures, the risk-free rate will be estimated as the average risk-free rate over the sample period. In estimating Jensen's, the risk-free rate will be allowed to vary over time following his procedures. Rate of return will be measured in two ways: the monthly investment relative, and the continuously compounded rate of return. The monthly investment relative can be defined as the wealth at the end of the month resulting from a one dollar investment at the beginning of the month with dividends reinvested. The continuously compounded rate of return is just the natural logarithm of the investment relative. Jensen argues that this is the appropriate measure of return if the market has an infinitesimal time horizon.

II. *Theoretical vs. Empirical Relationship of One-Parameter Performance Measures*

Theoretically, it would be expected that for random portfolios the Sharpe, Treynor, and Jensen one-parameter measures of performance would be independent of the corresponding measures of risk unless one or more of the following four conditions holds:

1) The assumptions underlying the market-line theory are invalid, i.e., are not realistic approximations of the real world, even for the *ex ante* magnitudes to which the theory applies.
2) The *ex post* distributions of return and values of risk differ substantially from their *ex ante* magnitudes.
3) Measurement errors, especially in the risk variables, result in biased estimates of the relationship between performance and risk.
4) There are in fact real systematic differences among the risk-adjusted performances of portfolios characterized by different degrees of risk.

The last possibility does not seem meaningful unless there are appreciable differences between *ex post* and *ex ante* magnitudes. If random portfolios do exhibit over very long periods of time significant dependencies between the one-parameter measures of performance and risk, the first and third conditions are more likely than the second to explain this result since *ex post* magnitudes would not be expected to deviate from their *ex ante* values indefinitely though such deviations are possible for extended periods.

The empirical analysis to follow suggests

mathematical discussion, the reader is referred to Jensen (1969) who indicates that the resulting bias is extremely small.

that the invalidity of a key assumption in the market-line theory does bias systematically the one-parameter measures of performance in all periods, while discrepancies between *ex post* and *ex ante* values (and perhaps also the invalidity of other assumptions) affect these measures of performance in different ways depending on the underlying market conditions. Measurement errors apparently do not substantially bias the estimates of the relationship between performance and risk.

To examine the relationship of one-parameter performance measures to risk, both performance and risk measures were derived for 200 random portfolios. These portfolios were selected from the universe of 788 common stocks listed on the New York Stock Exchange (*N.Y.S.E.*) throughout the period January 1960 through June 1968.[9] These 200 random portfolios consist of 50 individual portfolios of 25 securities and a like number for portfolios of 50, 75, and 100 stocks. An equal investment is assumed in each stock. A stratified random sampling procedure was used to insure that the entire spectrum of risk would be covered.[10]

Three basic performance measures computed for these 200 random portfolios were regressed against each of two measures of portfolio risk. The performance measures used were those developed by Sharpe, Treynor, and Jensen described in Section I. The two measures of risk were *Beta* coefficients, the covariance of portfolio and market return divided by the variance of market return, and the standard deviation of portfolio return. The first risk measure is implicit in the measures of Treynor and Jensen, and the second in Sharpe's. These regressions were derived using the performance and risk measures calculated with both the monthly investment relatives and the continuously compounded returns or the natural logarithms of the monthly relatives. Table 1 presents the resulting 12 regressions for the entire period January 1960 through June 1968. Scatter diagrams were also plotted for each of these 12 regressions, but only one is presented in view of space considerations (Figure 1).[11]

The results are striking. In all cases, risk-adjusted performance is dependent upon risk: The relationship is inverse and highly significant.[12] While rate of return is normally found to be positively related to risk, the adjustment of the rate of return for risk which would be expected to eliminate this relationship actually reverses it. For cross-sectional data, the correlations are quite high. The correlations for the *log relatives* are higher than the other correlations, with the differences particularly large for the Jensen regressions. The highest correlation is associated with the regression of Jensen's performance

[9] The monthly data for individual securities were obtained from University of Chicago updated tapes. These monthly measures of return reflect all capital gains as well as dividends and are adjusted for all capital changes (i.e., stock dividends, splits, etc.) as described in Lawrence Fisher and James Lorie. The market return refers to Fisher's combination link relatives, as updated for this study. These relatives assume an equal investment in each of the *N.Y.S.E.* stocks. The risk-free rate, also required to estimate the one-parameter performance measures, is the three-month yield on Treasury bills adjusted to a one-month basis.

[10] The *N.Y.S.E.* securities were ranked in ascending order according to the values of the covariances of the monthly security investment relatives and Fisher's link relatives over the entire sample period. A portfolio of 25 securities, one of 50, one of 75, and one of 100, were drawn at random from the first 200 securities. Four more portfolios of the four different sizes were drawn from the ranked 13th through 212th securities. This process was repeated again and again, each time increasing the bounds of the stratum by 12 securities until 200 portfolios were obtained.

[11] In view of the non-linearity observed in the scatter diagrams and the possible dependencies among portfolios, the regressions and t-values in Table 1 should be regarded as rough approximations.

[12] Section I contains formulae for the expected values of the constant terms in the theoretical relationships. The expected value is zero for the Jensen relationships and greater than zero for the other two.

TABLE 1—REGRESSIONS OF ONE-PARAMETER PERFORMANCE MEASURES ON RISK
Random Portfolios, January 1960–June 1968

Performance Measure	=	a	+	b Risk Measure	$\bar{R}^2$	Standard Error
1) S	=	0.2677 (45.50)	−	0.0557 X_1 (−9.17)	0.2944	0.0215
2) S	=	0.2724 (42.00)	−	1.3614 X_2 (−9.01)	0.2871	0.0216
3) T	=	0.0134 (45.10)	−	0.0039 X_1 (−12.82)	0.4510	0.0011
4) T	=	0.0136 (39.82)	−	0.0921 X_2 (−11.59)	0.4012	0.0011
5) J	=	0.0028 (11.34)	−	0.0018 X_1 (−7.13)	0.2004	0.0009
6) J	=	0.0029 (10.42)	−	0.0429 X_2 (−6.61)	0.1768	0.0009
7) S'	=	0.2648 (44.57)	−	0.0741 X_1' (−12.04)	0.4199	0.0214
8) S'	=	0.2714 (41.49)	−	1.8356 X_2' (−11.91)	0.4146	0.0215
9) T'	=	0.0130 (44.98)	−	0.0046 X_1' (−15.29)	0.5391	0.0010
10) T'	=	0.0133 (39.92)	−	0.1097 X_2' (−13.99)	0.4946	0.0011
11) J'	=	0.0345 (31.04)	−	0.0311 X_1' (−27.03)	0.7857	0.0040
12) J'	=	0.0372 (29.80)	−	0.7698 X_2' (−26.16)	0.7745	0.0041

Note: S represents the Sharpe measure of performance; T represents the Treynor measure of performance; J represents the Jensen measure of performance; X_1 represents the *Beta* coefficient of a random portfolio; X_2 represents the standard deviation of portfolio return; $\bar{R}^2$ represents the coefficient of determination adjusted for degrees of freedom.

The figures in parentheses are t-values. The unprimed variables are calculated using monthly relatives; the primed variables using the logarithm of the monthly relatives.

measure using *log relatives*, the variant preferred by Jensen, on the corresponding *Beta* coefficient. The large values of the coefficients of determination adjusted for degrees of freedom would suggest if these performance measures are valid that, at least for the period covered, the best way of ensuring good performance was to select a non-risky portfolio: In other words, risk is a good inverse proxy for performance. Not only is the proportion of variance in performance explained by risk very high in these relationships, but the implied magnitude of the impact of variations in risk on performance is sizable.

It has not proved feasible to extend our random portfolios back before 1960, but the performance measures for 115 mutual funds over the period 1945–64 presented by Jensen in the May 1968 issue of the *Journal of Finance* provide some useful insights into the earlier period. There have been many analyses indicating that, at least until recent years, the average performance of mutual funds in their stock investment has not deviated in any important way from that of random portfolios of *N.Y.S.E.* securities, although the funds have not confined themselves to investments in *N.Y.S.E.* issues.

The regression of the 1945–64 performance measures for these 115 funds on

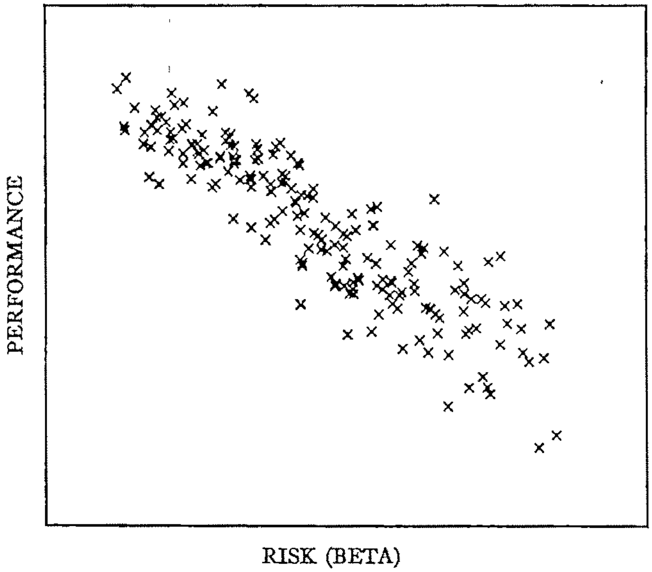


FIGURE 1. SCATTER DIAGRAM OF JENSEN'S PERFORMANCE MEASURE* ON RISK.
January 1960—June 1968
* Using *log relatives*.

the corresponding *Beta* coefficients[13] is

$$J' = .0399 - .0606X' \qquad \overline{R}^2 = .39$$
$$\quad (6.48) \quad (-8.57)$$

where J' is Jensen's performance measure using *log relatives* rather than the relatives themselves, X' is the associated *Beta* coefficient, the variables are estimated from annual rather than monthly data, the terms in parentheses are t-values, and $\overline{R}^2$ is the coefficient of determination adjusted for degrees of freedom.

Again the data over a long period of time indicate that performance and risk are strongly inversely correlated, with the riskiest portfolios performing very much worse than the less risky portfolios. Jensen (1969) shows that the performance funds have the highest *Beta* coefficients and the balanced funds the lowest coefficients, suggesting that these coefficients may serve as a reasonable proxy for risk. Neither Jensen, nor Donald Farrar who commented on Jensen's paper, made any reference to the apparent bias against the riskier funds over this period in the Jensen one-parameter measure of performance.

The first question that must be answered, in attempting to explain this apparent bias of one-parameter measures of performance against the riskier funds, is whether the bias is purely statistical in origin reflecting random errors in measurement of the risk variables which appear on both sides of the regression.[14] It can be shown, however, that if errors involved in measuring the risk variables are random (either with constant variance or with variance proportional to the square of risk), the estimate of the regression coefficient of the risk variable in the Jensen regression is biased downward, but the

[13] These coefficients were obtained from Jensen (1968, 1969).

[14] The risk variables appear in the denominator of the dependent variable and simultaneously as the independent variable in the Sharpe and Treynor regressions, and with opposite signs on each side of the Jensen regression.

magnitude of the bias is so small that it can be ignored.[15] Similarly, the estimate of the slope coefficient in the Sharpe and Treynor measures is asymptotically downward biased,[16] but again the likely magnitude of the bias is such that it could not account for the results in Table 1. Therefore, the assumptions underlying the market theory will be examined for further insights into the apparent bias in these performance measures.

[15] If the expected value of the coefficient of the risk variable in the Jensen regressions is b' and the estimated value is b, then it can be shown that the expected value of b

$$E(b)=\frac{b'-k\dfrac{\sigma_\epsilon^2}{\sigma_x^2}}{1+\dfrac{\sigma_\epsilon^2}{\sigma^2}}$$

where k is the difference between the rate of return on the market (R_m) and the riskless rate of return (R_f), ϵ is the measurement error in x, and x is the measure of risk. For the random portfolios using investment relatives, the variance of the measurement error, σ_ϵ^2, is estimated to be 0.0009883; σ_x^2 is 0.06198, and k is roughly .008. Therefore, if $b'=0$, the bias in its estimate is approximately -0.000125, which in absolute value is considerably less than the actual estimate -0.0557. For Jensen's sample of mutual funds, estimates of σ_ϵ^2 and σ_x^2 are, respectively, 0.0092 and 0.0413 and k is at most 0.16, so that if $b'=0$, the bias is approximately -0.02915 which is compared to the actual estimate of -0.0606. Thus, the bias for the random portfolios is trivial, and for Jensen's mutual funds can only explain about half of the actual estimate.

[16] For example, in the Treynor regressions, if $y_i=R_i-R_f$ and $x_i=\beta_i$, the risk coefficient, the correct relationship between x and y (dropping the subscripts) will be

$$\frac{y}{x}=a'+b'x+v$$

where a' and b' would be the regression coefficients without measurement error in x, and v is a random disturbance independent of x and with zero expectation. If, however, x is measured with random error ϵ, independent of x and with zero expectation, the estimated regression will be

$$\frac{y}{x+\epsilon}=a+b(x+\epsilon)+\nu$$

where a and b are the estimated coefficients and ν a random disturbance independent of x and ϵ and with zero expectation.

Then, if n is the number of observations, the estimated coefficient b will be

$$\frac{b'[\Sigma(x-\bar{x})^2]+n\bar{x}\left(\dfrac{\bar{y}}{x}-\dfrac{\bar{y}}{x+\epsilon}\right)-n\bar{\epsilon}\dfrac{\bar{y}}{x+\epsilon}}{\Sigma(x-\bar{x})^2+\Sigma(\epsilon-\bar{\epsilon})^2+2\Sigma(x-x)(\epsilon-\bar{\epsilon})}$$

In the limit, the last term in the numerator will be zero because the limit of a product of sequences is the product of the limits. The last term in the demoninator will be zero because x and ϵ are assumed independent.

The middle term in the numerator will probably be negative under the observed distribution for x and ϵ. (The estimated risk measures, $x+\epsilon$, were always positive and it seems reasonable to assume that the values of x are always positive.) If it is, assuming that the measurement error ϵ is symmetrically distributed about zero and in absolute value is less than the minimum value of x, which seems appropriate for the random portfolios, the following inequality holds for any x:

$$\frac{1}{x(x+\epsilon)}<\frac{1}{x(x-\epsilon)},\qquad \epsilon>0$$

Since (ϵ) and ($-\epsilon$) are equally likely,

$$\mathrm{E}\left(\frac{\epsilon}{x(x+\epsilon)}\,\Big|\,x\right)$$

will be negative for all x, so that

$$E\left(\frac{\epsilon}{x(x+\epsilon)}\,\Big|\,x\right)$$

$=h$ will be negative. Under the market-line theory, $E(y)=E(R_m)-R_f=k$ is greater than zero, which implies in the limit that

$$\bar{x}\left(\frac{\bar{y}}{x}-\frac{\bar{y}}{x+\epsilon}\right)=khE(x)<0$$

or that in the limit

$$b=\frac{b'+\dfrac{khE(x)}{\sigma_x^2}}{1+\dfrac{\sigma_\epsilon^2}{\sigma_x^2}},\qquad h<0$$

which yields a downward biased estimate of b'.

An estimate of the bias for the regression of Treynor's measure calculated with the investment relatives on the *Beta* coefficient (Table 1, equation 3) was derived. The expected value of x, $E(x)$, was estimated as 0.934. The estimates of σ_ϵ^2 and σ_x^2 are the same as those given in the preceding footnote. Assuming x and ϵ are normally distributed with means of 0.93 and 0.0 and with variances as given above, a simulation using 10,000 drawings yielded -0.00114 as an estimate of h. Thus, the bias is estimated as -0.000136, which should be compared to -0.0039, so that the bias is negligible. In view of the similarity between Treynor's and Sharpe's measure as well as the empirical results in Table 1, the relative bias in the Sharpe regressions is likely also to be trivial.

Of the key assumptions underlying the market theory leading to one-parameter measures of performance, the one which most clearly introduces a bias against risky portfolios is the assumption that the borrowing and lending rates are equal and the same for all investors. Since the borrowing rate for an investor is typically higher than the lending rate, the assumption of equality might be expected to bias the one-parameter measures of performance against risky portfolios because, for such portfolios, investors do not have the same option of increasing their return for given risk by moving from an all stock portfolio to an investment with additional stock financed with borrowings at the lending rate.

An examination of the scatter diagrams of the one-parameter performance measures against the risk measures confirms the strong inverse correlation between "risk-adjusted performance" and risk indicated by the regression in Table 1. The scatters show a fairly steady decline in these performance measures throughout the observed range of *Beta* coefficients with some evidence of a tapering in the rate of decline for the coefficients in excess of one (e.g., see Figure 1). The absolute values of the performance measures are in excess of market expectations for funds with *Beta* coefficients below one and below expectations for higher coefficients. These findings suggest that an "optimal" portfolio consisting of positive investments in both a risky portfolio and a riskless asset does not contain the market portfolio as its risky component, contrary to the usual assumption.[17] The risky portfolio involved in such optimal combinations appears to be considerably less risky than the market portfolio. If the difference between the borrowing and lending rates is sufficiently large, only risky assets would be held in portfolios with *Beta* coefficients beyond some low value. However, with more moderate differences between the borrowing and lending rates, both risky and riskless assets may be held, even in portfolios with high Beta coefficients.

Of the other departures from the perfect market assumptions of the market-line theory, none seems likely to introduce a substantial bias against risky portfolios. Differential taxation of capital gains and dividend income would tend to make a unit of before-tax return on risky portfolios larger on an after-tax basis than an equivalent before-tax return on less risky portfolios. The difference in after-tax return of the random portfolios for given before-tax return is likely to be quite small, however, since they are confined to *N.Y.S.E.* stocks. At a maximum, this difference between random portfolios with relatively small and those with relatively large *Beta* coefficients is likely to be not much over .3 of 1 percent annually,[18] and the difference would be negligible for portfolios with only moderate variations in these coefficients. Differential transaction and information costs might also affect moderately the comparison of performance of risky vs. less risky portfolios, but as discussed subsequently any bias which may exist from this source favors risky portfolios rather than the reverse.

[17] If borrowing and lending rates were equal, then it would be expected that the market line would be tangent to the efficient set at the point representing the market portfolio; but if the borrowing and lending rates are not equal, the market line (from the lending rate to the efficient set of risky securities) which represents optimal combinations of lending at the risk free rate and investing in a risky portfolio may touch the efficient set at a point representing higher or lower risk than that of the market portfolio.

[18] This estimate assumes that the average dividend yield on the risky portfolio is as much as 2 percent lower than on other portfolios which is offset by 2 percent more price appreciation; the former is assumed to be subject to a 15 percent higher tax rate than the latter. It should be pointed out that there is a significant though small group of corporate investors for whom the capital gains tax is higher than the tax on dividend income.

Still another limitation in the market-line theory which might help to explain at least part of the observed results is the assumption that it is possible to stipulate a holding period or planning horizon over which it is planned to hold all assets. If this is not possible, then risk-free assets may be desired for liquidity purposes[19] and their overall rate of return to the investor may be understated by the market rate. Such an understatement would tend to bias the one-parameter measures of performance downward for high risk portfolios (*Beta* coefficients higher than one) and upward for low risk portfolios. However, for realistic values of the understatement of the effective risk-free rate, the resulting biases would typically explain only a small part of the deviations of the one-parameter measures from their expected values.[20]

The only other explanation which comes to mind for the apparent bias against risky portfolios over the 1945–64 and 1960–68 periods is a difference between *ex ante* and *ex post* magnitudes, such that the *ex post* returns for risky portfolios in both periods were lower and the *ex post* risk higher than the respective *ex ante* values. However, as pointed out in Section I, Jensen has shown that if the risk measure *Beta* and the risk-free rate are constant over time, differences between *ex ante* expectations of returns and *ex post* realizations are irrelevant. The values of *Beta* for the random portfolios are remarkably constant over time,[21] and as shown by Jensen (1969), these same measures are reasonably stationary over time for his mutual funds. In addition, it seems intuitively highly unlikely that investors underestimated the risk of risky portfolios at the beginning of either of these periods since the whole evolution of the investment climate in the following years was a gradual realization that cyclical risks were no longer as great as they had been in our earlier history. Finally, the observed variations in the risk-free rate are likely to introduce only trivial biases.

Even though the risk measure *Beta*, which is really a measure of covariation with respect to general market movements, can be assumed to be stationary over time at least for the random portfolios, the performance measures may yet be biased because Jensen in his proof assumes that the *ex post* return on an individual security can be explained by two orthogonal factors: a market factor common to all securities and a unique factor. This is of course a simplistic view of the determination of returns. There are certainly industry factors, i.e., factors affecting a subset of securities, and possibly factors affecting stocks with different *Beta* coefficients. If the *ex post* values of these factors were such that the *ex post* return for portfolios with high *Beta* coefficients were lower than their *ex ante* values, the observed bias would result. However, the relationships between the average return on risky stocks and on other stocks in the 1945–64 and 1960–68 periods were somewhat more favorable to the risky stocks than in the preceding periods, 1926–45 and 1945–60, respectively.[22]

At this stage of our analysis, therefore, it appears that the unreality of the assumption of equal borrowing and lending rates is the most important factor explaining the bias of existing one-parameter measures of

[19] See Reuben Kessel.

[20] Thus, the bias in the Jensen measures resulting from an understatement of ΔR_f in the risk-free rate is $(1-\beta_i)\Delta R_f$ where β_i is the *Beta* coefficient for the i^{th} portfolio. At most, the value of ΔR_f might be on the order of 0.1 percent (per month). For $\Delta R_f=0.1$ percent, the bias for a portfolio with a low β_i say of 0.48 would be 0.00052, which might be compared with corresponding Jensen performance measure of 0.0022.

[21] The correlation between the *Betas* of the same portfolio for 1960–64 and 1964–68 was .96.

[22] These comparisons are based on data discussed in Friend and Paul Taubman. The data in that paper for the years 1926–60 have since been updated.

performance against risky portfolios over the extended periods of time analyzed.

Further insight into the relative importance of differences between borrowing and lending rates and of differences between *ex ante* and *ex post* magnitudes can be obtained by breaking down our longer period into shorter spans of time, segregating in particular those recent years when the market seemed to favor speculative issues to an unusual degree. We have broken down the period of January 1960 through June 1968 into two equal intervals: January 1960 through March 1964, and April 1964 through June 1968, the latter corresponding to a period of speculative fervor and booming prices of risky issues.

Tables 2 and 3 present regressions for each of the two sub-periods, January 1960 through March 1964, and April 1964 through June 1968, corresponding to those presented in Table 1 for the period as a whole. Figures 2 and 3 present the corresponding scatter diagrams. For the first of these two sub-periods, most of the same tendencies characterizing the period as a whole are observed but in even stronger form (Table 2 and Figure 2). The negative correlations between the performance and risk measures are extremely high, and the effects of the impact of variations in risk on performance, implied by the slope coefficients, are very substantial.[23] For the most recent period, however, the situation is reversed (Table 3 and Figure 3). The correlations between performance and risk measures are significantly positive and the slope coefficients fairly sizable though all eight slope coefficients and six of the eight correlations[24] are much lower than in the earlier period.

The most plausible explanation of the new results for April 1964 through June 1968 seems to be that in this interval the *ex post* returns for risky portfolios were higher than the respective *ex ante* values conditional on the general market factor[25] or that the *ex post* risks were lower than their *ex ante* values, and that these differences between *ex post* and *ex ante* magnitudes more than offset the normal bias operating in the opposite direction as a result of differences between borrowing and lending rates. There is a substantial body of evidence pointing to unanticipated high returns on risky issues in recent years as reflected in the much greater upsurge of price-earnings ratios on such securities than on the less risky issues. This upsurge in price-earnings ratios and hence in total return probably reflected a changed risk valuation of the riskier issues fully as much as higher than expected earnings. This changed risk valuation may have partially reflected a growing recognition both of the reduction in cyclical risks and of the potential for reducing risk through diversification. Whether such higher returns are likely to be retained in more normal periods might be questioned for reasons which we are developing in a forthcoming study, but this is not relevant to the subject of the present paper.[26]

[23] The correlations for the *log relatives* are no longer higher than the other correlations, and in the Jensen regressions the former sizable differences are reversed.

[24] The exceptions are the two Jensen regressions using the *log relatives*.

[25] Even though the values of *Beta* show little variation over time, *ex post* and *ex ante* values conditional on the market may very well differ if, as seems quite plausible, there are factors besides market and unique factors determining the returns of individual securities. This argument was developed above.

[26] Two possible additional reasons for the changed results from the earlier years to the more recent period should be mentioned: First, instead of a single market line there may exist segmented market lines for investors with greatly different tastes in risk, and a much higher proportion of investable funds may have been flowing in recent years to investors (e.g., the "performance" funds) with relatively little risk aversion. This explanation, which is perhaps less plausible than that previously presented, would be inconsistent with the assumption of identical or homogeneous expectations implicit in one-parameter performance measures and would again cast doubt on their validity.

Second, transactions costs might be expected to bias

Table 2—Regressions of One-Parameter Performance Measures on Risk
Random Portfolios, January 1960–March 1964

Performance Measure	=	a	+	b Risk Measure	$\bar{R}^2$	Standard Error
1) S	=	0.4008	−	0.2494 X_1	0.7534	0.0331
		(40.22)		(−24.68)		
2) S	=	0.4102	−	5.7307 X_2	0.7365	0.0342
		(37.98)		(−23.60)		
3) T	=	0.0188	−	0.0119 X_1	0.7478	0.0016
		(38.92)		(−24.31)		
4) T	=	0.0192	−	0.2722 X_2	0.7225	0.0017
		(36.10)		(−22.78)		
5) J	=	0.0105	−	0.0088 X_1	0.7145	0.0013
		(27.12)		(−22.34)		
6) J	=	0.0109	−	0.2020 X_2	0.7029	0.0013
		(26.22)		(−21.72)		
7) S'	=	0.4016	−	0.2727 X_1'	0.7847	0.0328
		(40.35)		(−26.95)		
8) S'	=	0.4123	−	6.3085 X_2'	0.7695	0.0339
		(38.19)		(−25.80)		
9) T'	=	0.0187	−	0.0128 X_1'	0.7787	0.0016
		(39.18)		(−26.48)		
10) T'	=	0.0191	−	0.2956 X_2'	0.7562	0.0016
		(36.45)		(−24.87)		
11) J'	=	0.0686	−	0.0389 X_1'	0.5397	0.0082
		(27.45)		(−15.31)		
12) J'	=	0.0702	−	0.9021 X_2'	0.5330	0.0083
		(26.63)		(−15.10)		

Note: See Table 1.

Table 3—Regressions of One-Parameter Performance Measures on Risk
Random Portfolios, April 1964–June 1968

Performance Measure	=	a	+	b Risk Measure	$\bar{R}^2$	Standard Error
1) S	=	0.1515	+	0.1328 X_1	0.4632	0.0384
		(15.75)		(13.14)		
2) S	=	0.1378	+	3.3747 X_2	0.4381	0.0393
		(12.36)		(12.50)		
3) T	=	0.0088	+	0.0036 X_1	0.2257	0.0018
		(19.73)		(7.68)		
4) T	=	0.0081	+	0.0981 X_2	0.2477	0.0018
		(16.36)		(8.16)		
5) J	=	−0.0031	+	0.0037 X_1	0.2693	0.0016
		(−7.53)		(8.62)		
6) J	=	−0.0038	+	0.1038 X_2	0.3057	0.0016
		(−8.45)		(9.41)		
7) S'	=	0.1434	+	0.1214 X_1'	0.4133	0.0381
		(14.75)		(11.88)		
8) S'	=	0.1317	+	3.1075 X_2'	0.3851	0.0390
		(11.68)		(11.21)		
9) T'	=	0.0081	+	0.0032 X_1'	0.1956	0.0017
		(18.64)		(7.03)		
10) T'	=	0.0076	+	0.0883 X_2'	0.2097	0.0017
		(15.48)		(7.34)		
11) J'	=	−0.0048	+	0.0154 X_1'	0.8615	0.0016
		(−11.50)		(35.19)		
12) J'	=	−0.0068	+	0.4086 X_2'	0.8626	0.0016
		(−14.52)		(35.36)		

Note: See Table 1.

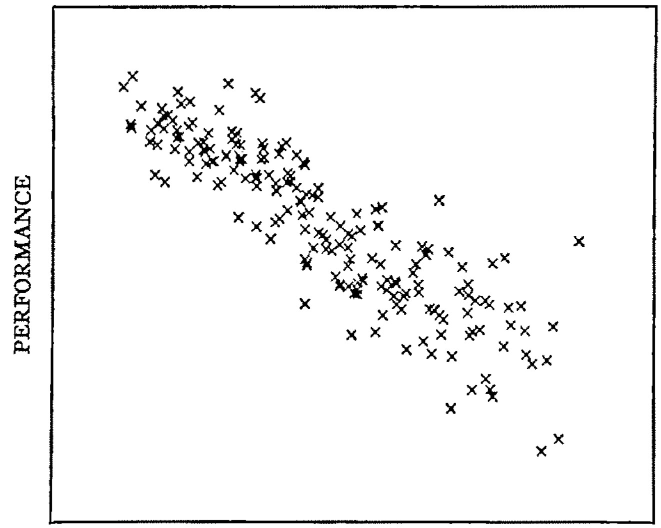


FIGURE 2. SCATTER DIAGRAM OF JENSEN'S PERFORMANCE MEASURE* ON RISK. January 1960—March 1964

* Using *log relatives.*

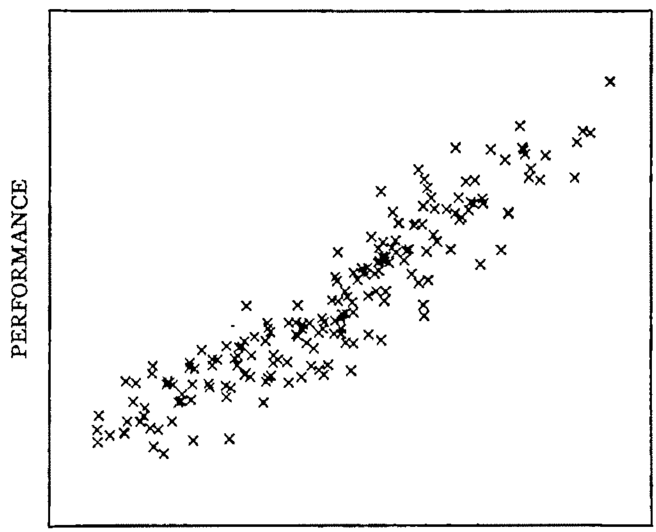


FIGURE 3. SCATTER DIAGRAM OF JENSEN'S PERFORMANCE MEASURE* ON RISK, April 1964—June 1968

* Using *log relatives.*

III. *Some Concluding Comments*

While the market-line theory has made a substantial contribution to our understanding of the theoretical mechanism of capital asset pricing under uncertainty, our analysis raises some questions about the usefulness of the theory in its present form to explain market behavior. The Sharpe, Treynor, and Jensen one-parameter measures of portfolio performance based on this theory seem to yield seriously biased estimates of performance, with the magnitudes of the bias related to portfolio risk. Thus, the numerous studies of mutual fund performance based on these one-parameter measures are suspect (e.g., Sharpe (1966), Jensen (1968), and Lintner (1965b)) especially when they attempt to appraise individual portfolios, or when the average risk of these portfolios differs from that of the market as a whole.

Somewhat improved measures of portfolio performance for any period could be obtained by adjusting the Sharpe, Treynor or Jensen measures of performance for the portfolio in question by the relationship between the corresponding measures of performance and risk of random portfolios in that same period, with the precise adjustment factor depending on the degree of risk in the non-random portfolio. However, in view of the remaining uncertainty as to the reasons for the observed biases, it is probably preferable at the present stage of knowledge to use the traditional two parameters—rate of return and risk—to measure portfolio performance, in preference to the more elegant but also more dangerous one-parameter measures, since in the former it is not necessary to stipulate an explicit functional relationship between risk and return.

the one-parameter performance measures somewhat in favor of the riskier portfolios. In estimating these measures on a monthly basis it is assumed that there is a risk-free investment (essentially a Treasury bill maturing in exactly one month) with a known return and no capital risk. For investors to take advantage of such instruments would require a turnover of this part of the portfolio every month, which would typically involve much higher turnover and probably somewhat higher relative costs than for the rest of the portfolio. The risk-free measure actually used in our empirical relationships is the three-month yield on Treasury bills adjusted to a one-month basis which in a period of markedly rising interest rates would probably introduce a small positive discrepancy on the average between actual and expected bill rates. This unanticipated increase in bill rates would bias the one-parameter measures of performance in favor of the riskier portfolios in the same manner as transaction costs, but the effect should be small.

REFERENCES

F. Arditti, "Risk and the Required Return on Equity," *J. Finance*, Mar. 1967, *22*, 19–36.

M. Blume, "Portfolio Theory: A Step Towards Its Practical Application," *J. Business*, Apr. 1970, *43*, 152-74.

E. Fama, "Risk, Return, and Equilibrium: Some Clarifying Comments," *J. Finance*, Mar. 1968, *23*, 29–40.

———, "Multiperiod Consumption-Investment Decisions," *Amer. Econ. Rev.*, Mar. 1970, *60*, 163–74.

D. Farrar, "Discussion: The Performance of Mutual Funds in the Period 1945–1964 by Michael C. Jensen," *J. Finance*, May 1968, *23*, 417–19.

L. Fisher, "Some New Stock-Market Indexes," *J. Business*, Jan. 1966, *39*, 191–225.

——— and J. Lorie, "Rates of Return on Investments in Common Stocks," *J. Business*, Jan. 1964, *37*, 1–21.

I. Friend and P. Taubman, "Risk and Stock Market Performance," *Proc. Center for Research on Security Prices*, Univ. Chicago November 1966.

M. Jensen, "The Performance of Mutual Funds in the Period 1945–1964," *J. Finance*, May 1968, *23*, 389–416.

———, "Risk, Capital Assets, and the Evaluation of Investment Portfolio," *J. Business*, Apr. 1969, *42*, 167–247.

R. Kessel, *The Cyclical Behavior of the Term Structure of Interest Rates*, New York 1965.

J. Lintner, (1965a) "The Valuation of Risk Assets and the Selection of Risky Investments

in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.*, Feb. 1965, *47*, 13–37.

———, (1965b) "Security Prices, Risk, and Maximal Gains from Diversification," *J. Finance*, Dec. 1965, *20*, 587–616.

H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, New York 1959.

W. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. Finance*, Sept. 1964, *19*, 425–42.

———, "Mutual Fund Performance," *J. Business*, Jan. 1966, *39*, 119–38.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. Econ. Stud.*, Feb. 1958, *25*, 65–85.

J. Treynor, "How to Rate Management of Investment Funds," *Harvard Bus. Rev.*, Jan.-Feb. 1965, *43*, 63–75.

———, "Toward a Theory of Market Value of Risky Assets," unpublished manuscript, undated.

Production and Investment in Natural Resource Industries

By OSCAR R. BURT AND RONALD G. CUMMINGS*

There has been a recent tendency for economists to view natural resource scarcity as a diminishing threat to man's existence and his ability to produce an abundant supply of worldly goods (see John V. Krutilla). This change in view has emanated from the world's experience of the last few decades with technological development, and the absence of any evidence that this rapid accumulation of technical knowledge will cease.

In some respects, one might think that modern technology has removed the need for an economic theory of conservation (temporal allocation) of natural resources. On the other hand, conservation of natural resources can now be looked upon with less emotion since the dire consequences of too rapid exploitation are no longer a threat. The question of conservation is thus a cold economic problem removed from many of its older value connotations.

There is indeed, considerable effort taking place in the economics profession to fill this need for a quantitative economic theory of conservation; however, with one exception (Vernon Smith), focus has been limited to *specific* resources. Earlier works by Lewis C. Gray and Harold Hotelling concerning the "Theory of the Mine" have recently been extended in numerous papers (see R. Cummings and O. Burt, Richard L. Gordon, Anthony Scott 1967). Problems associated with petroleum production are considered in Paul Davidson, and the fishery is considered in J. A. Crutchfield and Arnold Zellner, H. Scott Gordon, James Quirk and Vernon Smith, Anthony Scott (1955), and Ralph Turvey. An extensive literature exists on the temporal allocation of groundwater; two recent contributions are Gardner Brown and Charles B. McGuire, and Burt.[1]

A logical problem has been encountered, however, in efforts to examine the optimum allocation of a natural resource over time without explicitly considering the rate of investment associated with resource use. This difficulty stems from the long life of many capital items used in the process of exploiting the natural resource. Earlier studies have either tacitly or explicitly amortized costs of investment associated with resource use and proceeded as though all costs of production were essentially variable. Smith was the first to explicitly bring capital investment into a production model for natural resources, and his work was followed by the authors where only exhaustible resources were considered. More recently, Quirk and Smith have introduced a dynamic model of the fishery in which capital investment is treated explicitly.

It seems then that there is a need for a comprehensive model for simultaneously optimizing the rate of resource use and investment in natural resource industries

* The authors are, respectively, professor of agricultural economics and economics at Montana State University, and economist with Resources for the Future, Inc. This article is a contribution from Montana Agricultural Experiment Station, Journal Series No. 133.

[1] This list is by no means meant to be all inclusive concerning works related to natural resource use. No discussion of conservation economics would be complete without mention of the many contributions of S. V. Ciriacy-Wantrup.

in general. Such a model should be sufficiently general that the functional relations used could be adapted and the results applicable to any specific resource. Earlier reports concerning specific resources would fall out as a special case of a general theory concerning the intertemporal allocation of natural resources.

The purpose of this paper is to attempt to provide such a model. Optimization is viewed from the standpoint of society; however, our "social benefit" function may easily be viewed as a profit function and the results applied to the optimum behavior of the firm. The paper takes the following form: in Section I, a general model for social production and investment in natural resource industries is specified. In Section II, an economic interpretation of the characteristics for optimum time paths of production and investment are discussed under the assumption that society's planning horizon is finite. In Section III, equilibrium conditions under an infinite planning horizon are considered. Finally, in Section IV, our results are generalized to include multiple interrelated resources. The Appendix takes up some technical problems associated with the infinite planning horizon and second-order conditions for optimization.

I. *A General Production Model*

We begin our analysis with the following definitions and assumptions. Let t=0 and t=T be the beginning and end, respectively, of society's planning horizon. This period will be taken as finite, but later in our analysis we shall consider the case where $T \to \infty$. Define u_t as the rate at which the resource is used at time t, x_t as the amount of resource in stock at t, and y_t as capital stocks at t. The variable u_t is assumed subject to the constraints

$$(1) \quad 0 \leq u_t \leq h(x_t, y_t) \quad t = 0, \ldots, T-1$$

where $h(x_t, y_t)$ is assumed concave. The relationship $h(x_t, y_t)$ is such that larger resource and capital stocks allow for larger rates of production, i.e., its first partial derivatives are nonnegative.

In referring to u_t, we use the terms rate of production, rate of extraction, and rate of use as though they were synonymous. One or the other is most appropriate in a particular resource industry, but neither fits all situations. In some applications, the physical and biological nature of the resource may require u_t to be an indirect control. For example, an application to livestock range management is conveniently formulated with u_t defined as the number of animal units per acre, and the resource stock variable might be a composite index of range conditions (instead of a physical stock of an easily quantified natural resource such as groundwater). Another example is found in the discussion of the commercial fishery by Crutchfield and Zellner, where x_t is some measure of stocks of the fish specie in the sea and u_t is a measure of "fishing effort.'" However, u_t might be defined in the same units as x_t by using a fixed functional relationship between tons of fish harvested and the pair of variables, fishing effort and specie-stocks. At a given level of stocks, a specific fishing effort would imply a unique amount of harvested fish, and vice versa, so that the choice of a control variable is arbitrary.

At the beginning of any period t+1, resource and capital stocks are given by the difference equations,

$$(2) \quad x_{t+1} = x_t + g(u_t, x_t) \qquad x_0 = a$$

$$(3) \quad y_{t+1} = y_t - D(u_t, v_t, y_t) \qquad y_0 = b$$

$$t = 0, 1, \ldots, T-1$$

where v_t is gross addition to capital stocks at t. The only restriction on v_t is nonnegativity; no upper limit is specified. The functions $g(u_t, x_t)$ and $D(u_t, v_t, y_t)$ are assumed concave and convex, respectively.

The function $g(u_t, x_t)$ is net additions

to resource stocks during period t. The relationship $g(u_t, x_t)$ is such that (relatively) larger resource stocks and smaller rates of extraction result in greater additions to resource stocks, i.e., $\partial g/\partial x_t \geq 0$ and $\partial g/\partial u_t \leq 0$. The simplest case where u_t is in the same units as x_t would usually imply $\partial g/\partial u_t = -1$. In the case of groundwater, $g(u_t, x_t)$ is recharge to the groundwater aquifer minus the amount of water pumped. In a mineral industry, $g(u_t, x_t)$ is the difference between new discoveries and quantity extracted during the period t. For the fishery, u_t is usually defined as fishing effort instead of quantity of fish harvested, in which case, $g(u_t, x_t)$ reflects a combination of ecological and technological growth factors associated with the particular species of fish (see Smith).

The function $D(u_t, v_t, y_t)$ is a depreciation function for capital stocks which measures net changes during period t. A non-decreasing relation is assumed between depreciation and the rate of resource use as well as for the level of capital stocks, while a non-increasing relation is assumed for gross investment, i.e., $\partial D/\partial u_t \geq 0$, $\partial D/\partial y_t \geq 0$, $\partial D/\partial v_t \leq 0$. The simplest formulation would define v_t in the same units as y_t, giving $\partial D/\partial v_t = -1$, but it would be advantageous in some applications to have a more subtle relation between y_{t+1} and v_t. The presence of u_t in the depreciation function measures the consumption of capital associated with current production.[3]

If capital stocks, y_t, and capital investment, v_t, are limited in scalar measures, they must be interpreted as indices specific to a particular natural resource industry. At this point, we conceive them as such indices, but will later develop the case where the variables u_t, v_t, x_t, and y_t are vectors.

The functions $g(x_t, u_t)$ and $D(u_t, v_t, y_t)$ could be made explicitly dependent on time to allow for changing technical knowledge, but the same effect can be accomplished by defining capital measures specific to the technology at t=0, and taking changes in technology into account through the social benefit function (introduced below). Advancing technology would then be measured indirectly through fewer inputs used to achieve the same measure of capital investment in units of the index defined for t=0. This approach requires only that the benefit function be made explicitly a function of time.

It is convenient for later discussion to introduce a concave terminal value function for the natural resource and capital stocks, $\psi(x_T, y_T)$. It is difficult to conceive terminal stocks having value to society at the end of its planning horizon. Nevertheless, the function $\psi(x_T, y_T)$ is useful to analyze the implications of an infinite time-horizon.

We assume the existence of a concave social benefit function in the general form $B_t(u_t, v_t, x_t, y_t)$, and further assume that society wishes to maximize the present value of social benefits attributable to resource use during the time interval [0, T].

The rationale for including u_t in the benefit function is quite obvious since u_t is the rate of use of the resource which is assumed to have value in economic production, or as a direct consumption good. Likewise, v_t is a measure of periodic capital investment and has opportunity costs in the economy. The role of x_t and y_t is a little more subtle. Resource stocks, x_t, reflect accessibility of the resource. For example, large stocks of fish permit capture of a given quantity at a lower cost, and large stocks of a mineral are associated with relatively accessible and rich ores.

[3] The authors wish to express their appreciation to Vernon Smith for his comments on the form of the depreciation function during the presentation of an earlier draft of this paper at the University of Massachusetts.

$$(5) \qquad L = \sum_{t=0}^{T-1} \Big\{ B_t(u_t, v_t, x_t, y_t)\beta^t - \lambda_{t+1}\beta^{t+1}[x_{t+1} - x_t - g(u_t, x_t)] - \mu_{t+1}\beta^{t+1}[y_{t+1} - y_t + D(u_t, v_t, y_t)] - \nu_t\beta^t[u_t - h(x_t, y_t)] \Big\} - \lambda_0(x_0 - a) - \mu_0(y_0 - b) + \beta^T\psi(x_T, y_T),$$

Capital stocks, y_t, measure the amount of fixed capital invested in the natural resource industry, and thus higher stocks of capital imply lower costs of production per period.

All of the relationships involved are for a specific situation and the influence of the variables must be interpreted within that context. Relatively high resource stocks merely mean higher stocks at a particular point in time as compared to what the level might be under a different temporal allocation policy.

Let $\beta = 1/(1+r)$, where r is the appropriate positive discount rate. Since analysis of a single natural resource industry is a partial analysis relative to national economic growth, the discount rate should reflect the opportunity cost of capital to the remainder of the economy, and not a social discount rate per se.[8]

Our postulated societal optimization problem, vis-a-vis the use of a particular natural resource, is then given by

$$(4) \quad \text{Max}\left[\sum_{t=0}^{T-1} B_t(u_t, v_t, x_t, y_t)\beta^t + \beta^T\psi(x_T, y_T) \right]$$

subject to

a) $x_{t+1} - x_t - g(u_t, x_t) = 0$

b) $y_{t+1} - y_t + D(u_t, v_t, y_t) = 0$

c) $u_t - h(x_t, y_t) \leq 0,$

d) $u_t, v_t, x_t, y_t \geq 0$

e) $x_0 = a$ and $y_0 = b$

where the variables of optimization are u_t, v_t, $t=0, 1, \ldots, T-1$, and x_t, y_t, $t=1, \ldots, T$.

We assume that the functions in (4) are continuous and possess continuous first partial derivatives with respect to each argument. Following Kuhn and Tucker we form the Lagrangian Expression (5) where the multipliers, λ_t, μ_t, and ν_t are written with a discounting factor for the purpose of economic interpretation. This is also the purpose of including the terms associated with λ_0 and μ_0 which are, in a sense, redundant for the optimization problem.

As a notational device, the subscript t on the functions g, D, and h in the partial derivative of these functions implies that the partial derivative is evaluated with its arguments set equal to their magnitudes in period t. For example, $\partial g_t/\partial x_t$ is the partial derivative of $g(u, x)$ with respect to x, evaluated at $u = u_t$ and $x = x_t$.

Partial derivatives of L are listed below for future reference:

$$(6) \quad \frac{\partial L}{\partial u_t} = \left[\frac{\partial B_t}{\partial u_t} + \beta\lambda_{t+1}\frac{\partial g_t}{\partial u_t} - \beta\mu_{t+1}\frac{\partial D_t}{\partial u_t} - \nu_t\right]\beta^t,$$

$$t = 0, 1, \ldots, T-1$$

[8] For an exceptional treatment of the fishery resource within the framework of General Equilibrium Theory, see Quirk and Smith.

$$(7)\quad \frac{\partial L}{\partial v_t} = \left[\frac{\partial B_t}{\partial v_t} - \beta\mu_{t+1}\frac{\partial D_t}{\partial v_t}\right]\beta^t, \qquad t = 0, 1, \ldots, T-1$$

$$(8)\quad \frac{\partial L}{\partial x_0} = \frac{\partial B_0}{\partial x_0} + \beta\lambda_1\left(1 + \frac{\partial g_0}{\partial x_0}\right) + \nu_0\frac{\partial h_0}{\partial x_0} - \lambda_0$$

$$\frac{\partial L}{\partial x_t} = \left[\frac{\partial B_t}{\partial x_t} - \lambda_t + \beta\lambda_{t+1}\left(1 + \frac{\partial g_t}{\partial x_t}\right) + \nu_t\frac{\partial h_t}{\partial x_t}\right]\beta^t, \qquad t = 1, 2, \ldots, T-1$$

$$\frac{\partial L}{\partial x_T} = \left[\frac{\partial \psi}{\partial x_T} - \lambda_T\right]\beta^T$$

$$(9)\quad \frac{\partial L}{\partial y_0} = \frac{\partial B_0}{\partial y_0} + \beta\mu_1\left(1 - \frac{\partial D_0}{\partial y_0}\right) + \nu_0\frac{\partial h_0}{\partial y_0} - \mu_0$$

$$\frac{\partial L}{\partial y_t} = \left[\frac{\partial B_t}{\partial y_t} - \mu_t + \beta\mu_{t+1}\left(1 - \frac{\partial D_t}{\partial y_t}\right) + \nu_t\frac{\partial h_t}{\partial y_t}\right]\beta^t, \qquad t = 1, 2, \ldots, T-1$$

$$\frac{\partial L}{\partial y_T} = \left[\frac{\partial \psi}{\partial y_T} - \mu_T\right]\beta^T$$

We wish to derive values for the multipliers λ_t and μ_t by an iterative process explained below. In order to do this, we must assume that x_t and y_t will be positive for all t, $0 \leq t \leq T$. With but rare exceptions, it seems plausible to expect that stocks of the natural resource will indeed remain at positive levels throughout the period of economic production. In some cases, zero stocks would make it impossible to produce at positive levels of production again; biological resources such as fish and wildlife are examples.

It also seems that capital stocks would be at positive levels throughout the period of economic production. Capital stocks may be zero initially and for some time before production begins, but once production is started, one would expect stocks of capital to remain positive until the resource industry is abandoned. At any rate, we can think of a cycle of production where both natural resources and capital stocks are positive on an interval [0, T], and the terminal value function $\psi(x_T, y_T)$ reflects the future value of the industry for $t > T$.

Necessary conditions for a maximum in (4) are that (6) through (9) be nonpositive. If, at any t, a variable of optimization is positive, the associated partial derivative of L must then be zero. In view of the above discussion on positive levels of natural resource and capital stocks, we set the partial derivatives in (8) and (9) equal to zero as necessary conditions for maximum discounted social benefits within a production cycle.

The result for (8) is a recurrence relation in λ_t,

$$(10)\quad \lambda_t = \frac{\partial B_t}{\partial x_t} + \nu_t\frac{\partial h_t}{\partial x_t} + \beta\lambda_{t+1}\left(1 + \frac{\partial g_t}{\partial x_t}\right)$$

Iterating the above recursion for t, $t+1, \ldots, T$ yields equation (11). A comparative derivation from (9) gives equation (12).

At this point, we make a short digression to obtain simplifications in (11) and (12). Differentiating equations (2) and (3) we get

$$(13)\quad \frac{\partial x_{t+1}}{\partial x_t} = 1 + \frac{\partial g_t}{\partial x_t}$$

$$\frac{\partial y_{t+1}}{\partial y_t} = 1 - \frac{\partial D_t}{\partial y_t}$$

$$(11)\quad \lambda_t = \frac{\partial B_t}{\partial x_t} + \nu_t \frac{\partial h_t}{\partial x_t} + \beta\left[\frac{\partial B_{t+1}}{\partial x_{t+1}} + \nu_{t+1}\frac{\partial h_{t+1}}{\partial x_{t+1}} + \beta\lambda_{t+2}\left(1 + \frac{\partial g_{t+1}}{\partial x_{t+1}}\right)\right]\left(1 + \frac{\partial g_t}{\partial x_t}\right)$$

$$\vdots$$

$$= \left(\frac{\partial B_t}{\partial x_t} + \nu_t \frac{\partial h_t}{\partial x_t}\right) + \sum_{j=t+1}^{T-1}\left\{\left(\frac{\partial B_j}{\partial x_j} + \nu_i \frac{\partial g_i}{\partial x_j}\right)\prod_{i=t}^{j-1}\beta\left(1 + \frac{\partial g_i}{\partial x_i}\right)\right\}$$

$$+ \frac{\partial \psi}{\partial x_T}\prod_{i=t}^{T-1}\beta\left(1 + \frac{\partial g_i}{\partial x_i}\right)$$

$$(12)\quad \mu_t = \left(\frac{\partial B_t}{\partial y_t} + \nu_t \frac{\partial h_t}{\partial y_t}\right) + \sum_{j=t+1}^{T-1}\left[\left(\frac{\partial B_j}{\partial y_j} + \nu_i \frac{\partial h_j}{\partial y_i}\right)\prod_{i=t}^{j-1}\beta\left(1 + \frac{\partial D_i}{\partial y_i}\right)\right]$$

$$+ \frac{\partial \psi}{\partial y_T}\prod_{i=t}^{T-1}\beta\left(1 - \frac{\partial D_i}{\partial y_i}\right)$$

$$(15)\quad \lambda_t = \sum_{j=t}^{T-1}\beta^{j-t}\left(\frac{\partial B_j}{\partial x_j} + \nu_j \frac{\partial h_j}{\partial x_j}\right)\frac{\partial x_j}{\partial x_t} + \beta^{T-t}\frac{\partial \psi}{\partial x_T}\frac{\partial x_T}{\partial x_t},$$

$$(16)\quad \mu_t = \sum_{j=t}^{T-1}\beta^{j-t}\left(\frac{\partial B_j}{\partial y_j} + \nu_j \frac{\partial h_j}{\partial y_i}\right)\frac{\partial y_j}{\partial y_t} + \beta^{T-t}\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_t}, \qquad t = 1, 2, \ldots, T.$$

$$(17)\quad \frac{\partial B_t}{\partial u_t} - \nu_t \leqq \left[\sum_{j=t+1}^{T-1}\beta^{j-t}\left(\frac{\partial B_j}{\partial x_j} + \nu_j \frac{\partial h_j}{\partial x_j}\right)\frac{\partial x_j}{\partial x_{t+1}} + \beta^{T-t}\frac{\partial \psi}{\partial x_T}\frac{\partial x_T}{\partial x_{t+1}}\right]\left(-\frac{\partial g_t}{\partial u_t}\right)$$

$$+ \left[\sum_{j=t+1}^{T-1}\beta^{j-t}\left(\frac{\partial B_j}{\partial y_j} + \nu_j \frac{\partial h_j}{\partial y_j}\right)\frac{\partial y_j}{\partial y_{t+1}} + \beta^{T-t}\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_{t+1}}\right]\left(\frac{\partial D_t}{\partial u_t}\right)$$

$$(18)\quad \frac{\partial B_t}{\partial v_t} \leqq \left[\sum_{j=t+1}^{T-1}\beta^{j-t}\left(\frac{\partial B_j}{\partial y_j} + \nu_j \frac{\partial h_j}{\partial y_j}\right)\frac{\partial y_j}{\partial y_{t+1}} + \beta^{T-t}\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_{t+1}}\right]\left(\frac{\partial D_t}{\partial v_t}\right)$$

Successively applying the chain rule for implicit functions,

$$(14)\quad \frac{\partial x_j}{\partial x_t} = \prod_{i=t}^{j-1}\left(1 + \frac{\partial g_i}{\partial x_i}\right), \qquad j > t$$

$$\frac{\partial y_j}{\partial y_t} = \prod_{i=t}^{j-1}\left(1 - \frac{\partial D_i}{\partial y_i}\right), \qquad j > t$$

Substitution of (14) into (11) and (12) gives the simplified and more intuitive equations[4] (15) and (16).

Substituting (15) and (16), evaluated at t+1, into (6), (16) into (7), and applying the Kuhn-Tucker theorem, we find that necessary conditions for a maximum in (4) are (17) and (18). For u_t or v_t positive, relation (17) or (18) is a strict equality. We also note that ν_t is positive or zero as $u_t = h(x_t, y_t)$ or $u_t < h(x_t, y_t)$.

[4] For (15) and (16) to equal (11) and (12), the following notational adjustment must be made:

$$\prod_{i=t}^{t-1} a_i = 0$$

$$(19)\qquad \frac{\partial B_0}{\partial u_0} - \nu_0 = \beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right) + \beta\mu_1\frac{\partial D_0}{\partial u_0}$$

$$= \left[\sum_{j=1}^{T-1}\beta^j\left(\frac{\partial B_j}{\partial x_j} + \nu_j\frac{\partial h_j}{\partial x_j}\right)\frac{\partial x_j}{\partial x_1} + \beta^T\frac{\partial \psi}{\partial x_T}\frac{\partial x_T}{\partial x_1}\right]\left(-\frac{\partial g_0}{\partial u_0}\right)$$

$$+ \left[\sum_{j=1}^{T-1}\beta^j\left(\frac{\partial B_i}{\partial y_j} + \nu_j\frac{\partial h_j}{\partial y_j}\right)\frac{\partial y_j}{\partial y_1} + \beta^T\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_1}\right]\left(\frac{\partial D_0}{\partial u_0}\right)$$

$$(20)\qquad \frac{\partial B_0}{\partial v_0} = \beta\mu_1\frac{\partial D_0}{\partial v_0} = \left[\sum_{j=1}^{T-1}\beta^j\left(\frac{\partial B_j}{\partial y_j} + \nu_j\frac{\partial h_j}{\partial y_j}\right)\frac{\partial y_j}{\partial y_1} + \beta^T\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_1}\right]\left(\frac{\partial D_0}{\partial v_0}\right)$$

An immediate conclusion from (17) and (18) is that given x_t and y_t, the optimal levels of u_t and v_t are independent of u_i and v_i $i<t$. This is essentially a Markovian dependence structure in the decision process (see R. Bellman 1961, p. 54). The resource and capital stock variables completely summarize the impact of all previous decisions upon current and future optimal decisions.

In view of this recursive structure for optimal allocation through time, nothing is lost in generality by focusing on $t=0$ and thinking of any T-period allocation problem with x_0 and y_0 specified. Solution of (8) and (9) for λ_0 and μ_0 shows them to fit (15) and (16) for $t=0$, which justifies the previously questionable notation for these two multipliers associated with initial stocks.

Assume that u_0 and v_0 are positive at $t=0$. Then by (17) and (18), the optimum rates of production and investment are given by equations (19) and (20). Equations (19) and (20) will serve as the basis for the economic interpretations that follow.

II. *Economic Interpretations*

Attention is now focused on an interpretation of the results generated in Section I. While the solutions for u_t and v_t must be determined simultaneously within the model, we shall discuss the characteristics of optimal rates of production and investment separately.

The Optimum Time Path of Resource Use

At $t=0$, the optimum time path for production from the natural resource is given by (19). On the left-hand side of (19) is the measure of marginal social benefits attributable to an increment in resource use at $t=0$, less the imputed value of the constraint (4-c) *if* strict equality should hold at $t=0$ $\{\nu_0=0 \text{ if } u_0<h(x_0,y_0)\}$.

On the right-hand side of (19), the partial derivatives $\partial g_0/\partial u_0$ and $\partial D_0/\partial u_0$ measure the marginal influence of u_0 on resource and capital stocks. Most applications imply $\partial g_0/\partial u_0<0$, and $\partial D_0/\partial u_0=-1$ when resource use is measured in the same units as resource stocks. The latter derivative, $\partial D_0/\partial u_0$, is marginal capital consumption from current production.

The factor $\beta\lambda_1$ is the discounted value of increments to resource stocks in all future periods associated with an increment to stocks in period $t=1$. Likewise, $\beta\mu_1$ is the discounted value of increments to capital stocks in future periods associated with an increment to capital stocks in period one. The right-hand side of (19) is essentially

the "user cost" reflecting " . . . the present value of future profit (social benefits) foregone by a decision to produce a unit of output today (at t=0)" (Scott 1967, p. 42) and (19) may be interpreted as the decision rule: increase production at t=0 until net marginal social benefits at t=0 (net of the "boundary cost" μ_0) equal user costs.

This economic interpretation is clarified by analyzing the separate components of $\beta\lambda_1$, as written in (19). An almost identical detailed explanation applies to $\beta\mu_1$, except that the stocks would be capital instead of the natural resource.

The term $\mu_j(\partial h_j/\partial x_j)$ is the implicit value of a marginal unit of the resource available in period j to relax the constraint $u_j \leq h(x_j,y_j)$. Thus, the sum $\partial B_j/\partial x_j + \mu_j(\partial h_j/\partial x_j)$ is the direct marginal value of the resource stock during period j, i.e., the increment to benefits in period j associated with an increment to resource stocks in the same period, *ceteris paribus*. Of course, in the terminal period, t=T, this same measure is $\partial\psi/\partial x_T$. Since no production is permitted by the model in that period, only a lump sum value is assigned to remaining stocks. These marginal values multiplied by $\partial x_j/\partial x_1$ are values which are comparable to an increment to stocks in period one instead of period j. This follows from the implicit function theorem of the calculus.

In summary, the left-hand side of (19) is marginal social value of current production and the right-hand side is our user cost consisting of two terms, 1) the discounted marginal value of a unit of resource retained in stocks instead of used in current production and 2) the discounted marginal value of capital stocks consumed by an increment to current production.

The Optimum Time Path of Social Investment

Equation (20) is most easily interpreted after multiplication by (-1). Then the left-hand side is marginal social cost of investment in the current period, while the right-hand side is the discounted value of the increment to all future capital stocks associated with an increment to current investment. The economic interpretation of the right-hand side of (20) is identical to the second cost term in (19) with the exception that the weight is $\partial D_0/\partial v_0$, the effect on current rates of depreciation of an increment in gross investment, rather than $\partial D_0/\partial u_0$.

Further Ramifications of the Model

Another interesting result is derived by solving for $\beta\mu_1$ in (20) and substituting it into (19) to get

$$(21)\quad \frac{\partial B_0}{\partial u_0} - \nu_0 = \beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right) - \frac{\partial B_0}{\partial v_0}\frac{\partial v_0}{\partial u_0}$$

The last term in (21) measures the capital cost associated with an increment to current production. By holding y_1 constant, total adjustment in capital at the margin must take place through variation in v_0 in order that the net effect on depreciation will be zero. Thus, $-\partial B_0/\partial v_0$ is marginal social cost with respect to current investment, and multiplication by $\partial v_0/\partial u_0$ converts it to marginal social cost with respect to current resource use insofar as associated capital consumption is concerned.

Going back to equation (14), we see that

$$(22)\quad \beta^t\frac{\partial x_t}{\partial x_1} = \beta\prod_{i=1}^{t-1}\left(1+\frac{\partial g_i}{\partial x_i}\right)/(1+r)$$

$$(23)\quad \beta^t\frac{\partial y_t}{\partial y_1} = \beta\prod_{i=1}^{t-1}\left(1-\frac{\partial D_i}{\partial y_i}\right)/(1+r)$$

The left-hand sides of (22) and (23) represent the discount factors on $(\partial B_t/\partial x_t + \nu_t\partial h_t/\partial x_t)$ and $(\partial B_t/\partial y_t + \nu_t\partial h_t/\partial y_t)$ appearing in the summations of (19) and (20).

The term $\partial g_i/\partial x_i$ measures the effect of

an incremental change in resource stocks on the rate of growth in resource stocks at period i. If resource use at $t=0$ results in lower resource stocks at $t=1$, then the product $\prod_{i=1}^{t-1}[1+(\partial g_i/\partial x_i)]$ reflects the fact that the losses in resource stocks at $t=1$ affect the rates of growth in the resource stock in *all* future time periods. Thus, $[1+(\partial g_i/\partial x_i)]$ may be viewed as an ecological or biological (in some cases) compounding factor. Therefore, (22) essentially represents the ratio of a compounding to a discounting factor, the former relating to productivity of the resource stock in terms of generating *additions* to resource stocks, and the latter reflecting the productivity of social capital in all lines of economic activity. If $\partial g_t/\partial x_t < r$, the net effect is a discounting in the user cost equation.

Looking at (23), the term $\partial D_i/\partial y_i$ is the marginal rate of capital decay representing the effect on the depreciation of capital stocks of an incremental change in the magnitude of capital stocks at $t=i$. The term $(1-\partial D_i/\partial y_i)$ may then be interpreted as an internalized (to the resource industry from society's point of view) marginal rate of capital survival. Thus, investment at $t=0$ yields benefits in all future periods which are discounted by the product of the social discount factor and a discount rate emanating from marginal effects on capital decay. The larger is $\partial D/\partial y$, the more severe is this discounting, and the rate of social investment in the natural resource industry at $t=0$ is diminished. Of course, as the marginal rate of capital decay in the resource industry becomes small, the rate of social investment in the natural resource industry is increased.

III. *Equilibrium Resource and Capital Stocks*

In this section, the implications of equilibrium stocks of the natural resource and capital are pursued. An infinite planning horizon is assumed with the social benefit function invariant through time. Essentially, the assumptions of (A−1) in the Appendix are made, although they seem stronger than is probably necessary. Under these assumptions (combined with earlier specifications of continuity and convexity), either an equilibrium must exist, or resource stocks go to zero in a finite period. An equilibrium of zero is not ruled out. Some of the technical problems associated with an infinite planning horizon are relegated to the Appendix.

The case where resource stocks go to zero in a finite period can be interpreted as a situation where discounting economically overrides maintenance of the resource stock. Of course, exhaustible resources are quite likely to fall in this category, and one would expect that the case where stocks asymptotically approach zero (the only equilibrium possible) would be the exception. From this point, we assume existence of an equilibrium and analyze its implications.

Equilibrium implies $x_t = x_{t-1} = \ldots$ and $y_t = y_{t-1} = \ldots$, which in turn implies

$$(24)\qquad \begin{aligned} g(u_t, x_t) &= 0, \\ D(u_t, v_t, y_t) &= 0, \qquad \text{for all } t \end{aligned}$$

But for (24) to hold, $u_t = u_{t+1} = \ldots$ and $v_t = v_{t+1} = \ldots$.

Let us now examine the implications of constancy of these variables on the necessary conditions for maximum discounted social benefit, viz., equations (19) and (20). We drop subscripts on the variables to denote independence through time. The functions appearing in (19) and (20) are then constant over t because all the arguments are fixed at constant values u, v, x, y. Likewise, the Lagrange multipliers υ, λ, and μ will be constant over time t. The multiplier υ will very likely be zero except in the special case where $u=h(x, y)$ at equilibrium.

$$
\begin{aligned}
(26)\qquad \frac{\partial B}{\partial u} - \nu &= \Bigg[\sum_{j=1}^{\infty} \beta^j \left(\frac{\partial B}{\partial x} + \nu \frac{\partial h}{\partial x} \right) \left(1 + \frac{\partial g}{\partial x} \right)^{j-1} \\
&\quad + \lim_{T \to \infty} \beta^T \frac{\partial \psi}{\partial x} \left(1 + \frac{\partial g}{\partial x} \right)^{T-1} \Bigg] \left(- \frac{\partial g}{\partial u} \right) \\
&\quad + \Bigg[\sum_{j=1}^{\infty} \beta^j \left(\frac{\partial B}{\partial y} + \nu \frac{\partial h}{\partial y} \right) \left(1 - \frac{\partial D}{\partial y} \right)^{j-1} \\
&\quad + \lim_{T \to \infty} \beta^T \frac{\partial \psi}{\partial y} \left(1 - \frac{\partial D}{\partial y} \right)^{T-1} \Bigg] \left(\frac{\partial D}{\partial u} \right) \\
&= \beta \left(\frac{\partial B}{\partial x} + \nu \frac{\partial h}{\partial x} \right) \left[1 - \beta \left(1 + \frac{\partial g}{\partial x} \right) \right]^{-1} \left(- \frac{\partial g}{\partial u} \right) \\
&\quad + \beta \left(\frac{\partial B}{\partial y} + \nu \frac{\partial h}{\partial y} \right) \left[1 - \beta \left(1 - \frac{\partial D}{\partial y} \right) \right]^{-1} \left(\frac{\partial D}{\partial u} \right)
\end{aligned}
$$

With this prelude, we start simplifying equations (19) and (20) under the assumption that the summations are convergent as $T \to \infty$. Since $g(u, x)$ and $D(u, v, x)$ are constant through time in equilibrium (their arguments are at fixed values), the partial derivatives of (14) at $t=1$ can be written

$$
\begin{aligned}
(25)\qquad \frac{\partial x_j}{\partial x_1} &= \prod_{i=1}^{j-1} \left(1 + \frac{\partial g}{\partial x} \right) = \left(1 + \frac{\partial g}{\partial x} \right)^{j-1} \\
\frac{\partial y_i}{\partial y_1} &\quad \prod_{i=1}^{j-1} \left(1 - \frac{\partial D}{\partial y} \right) = \left(1 - \frac{\partial D}{\partial y} \right)^{j-1}
\end{aligned}
$$

Substitution of (25) into (19), and letting $T \to \infty$, we have equation (26). The corresponding derivation in (20) yields

$$
(27)\qquad \frac{\partial B}{\partial v} = \beta \left(\frac{\partial B}{\partial y} + \nu \frac{\partial h}{\partial y} \right) \cdot \left[1 - \beta \left(1 - \frac{\partial D}{\partial y} \right) \right]^{-1} \left(\frac{\partial D}{\partial v} \right)
$$

Convergence of the infinite sums requires that[5]

$$
\begin{aligned}
(28)\qquad &|1 + \partial g/\partial x| < 1 + r \\
&|1 - \partial D/\partial y| < 1 + r
\end{aligned}
$$

Assumptions (A-1-b) and (A-1-c) indirectly force these conditions, but a direct examination of the implications of (28) is instructive. We would expect $\partial g/\partial x > 0$ throughout the economically relevant range of resource stocks. If the resource is fish or wildlife, $\partial g/\partial x < 0$ would imply such large populations that an increment to the population causes a net decrease for next year (starvation, disease hazards, etc., could be the cause). Logically, $\partial D/\partial y > 0$ and $|\partial D/\partial y| < 1$ since $D(u, v, y)$ is a depreciation function and y is an index of capital stocks. It is hard to imagine a case where an increment to capital stocks would be associated with a

[5] This result is similar to that obtained by Quirk and Smith. In their dynamic model of the fishery, they show that necessary and sufficient conditions for user costs to be positive (under an infinite planning horizon) is that the productivity of the fish stock in producing *additional* stocks be less than the social rate of discount (p. 16).

decrement in stocks a year later that exceeded the initial increment, *ceteris paribus*. The one reasonable possibility that could cause divergence is $\partial g/\partial x > r$, but this case also is extremely unlikely since the natural resource stocks would ultimately reach such high levels that natural phenomena would limit their unbounded growth, and ultimately $\partial g/\partial x$ would drop below r. Therefore, assumptions (A-1-b) and (A-1-c) seem quite plausible for any conceivable application.

Economic interpretation of (26) and (27) is the same as for (19) and (20). The left side of (26) is marginal social benefits with respect to resource use in the current period, while the right-hand side is the sum of two terms: 1) "capitalized value" of an increment to resource stocks multiplied by the marginal effect that current production has on next year's stocks and 2) capitalized value of an increment to capital stocks multiplied by the marginal effect that current production has on next year's capital stocks. We are using capitalized value to describe reduction of a perpetual flow to a present value measure when the "discount factor" is $\beta(1+\partial g/\partial x)$ or $\beta(1-\partial D/\partial y)$. These latter discount factors are a combination of an economic factor, β, and a stock compounding (discounting) factor, $1+\partial g/\partial x$ (or $1-\partial D/\partial y$). The analogous interpretation for (27) is obvious.

Dividing (27) by $\partial D/\partial v$ and making a substitution into (26), we get the equilibrium counterpart of (21), i.e.,

$$(29)\quad \frac{\partial B}{\partial u} - \nu = \beta\left(\frac{\partial B}{\partial x} + \nu\frac{\partial h}{\partial x}\right) \cdot \left[1 - \beta\left(1 + \frac{\partial g}{\partial x}\right)\right]^{-1}\left(-\frac{\partial g}{\partial w}\right) - \frac{\partial B}{\partial v}\,\frac{\partial v}{\partial u}$$

The last term is an alternative expression for the implicit cost in capital consumption associated with an increment to resource use in the current period. This cost is expressed as the cost of current investment required to maintain capital stocks, while in (26) this same cost is expressed as the implicit cost of having incrementally less capital stock in future periods.

If the constraint, $u \leq h(x, y)$, is not binding at equilibrium, solution of the system of four equations, (24), (26), and (27) in the variables u, v, x, and y will provide the equilibrium solution. Should the upper constraint on u be binding $u = h(x, y)$ is a fifth equation and v is the fifth variable of equilibrium. This straightforward solution to determine equilibrium stocks and rates of resource use and investment is attractive for easy analysis in applications. A good deal can be learned about optimal production and investment in a resource industry from knowledge of equilibrium under an optimal policy.

IV. *A Generalization to Multiple Interrelated Resources*

The preceding analysis is now generalized to N, interrelated natural resources and M, interrelated indices of capital stocks. Fortunately, all of the preceding analysis is applicable after defining certain vector and matrix relationships. The functions $B_t(u_t, v_t, x_t, y_t)$ and $\psi(x_T, y_T)$ are kept as scalars, while $g(u_t, x_t)$ and $h(x_t, y_t)$ are vectors of N components each and $D(u_t, v_t, y_t)$ has M components. The variables u_t and x_t are N component vectors, and v_t and y_t are vectors of M components each. The above vectors are all column vectors.

In the Lagrangian relationship, the multipliers λ_t, μ_t, and v_t are defined as row vectors of N, M, and N components. These definitions make the partial derivatives of L in equations (6) through (9) row vectors, and also imply that the partial derivatives of B_t and ψ are row vectors. The partial derviatives of D, g, and h are matrices with rows equal in number to the dimension of

the vector function, and number of columns equal to the number of components in the vector variables of differentiation. Also, in expressions like $(1+\partial g_t/\partial x_t)$, the scalar 1 is replaced by the identity matrix I. With these conventions, all the derivations and equations presented for scalar relationships are valid. Ordering of factors in multiplication were kept consistent with the vector-matrix generalization, and divisions were even indicated by the inverse notation for easy transition.

In the derivation of equation (26), there arises the infinite sum

$$\sum_{j=0}^{\infty} \beta^j \left(I + \frac{\partial g}{\partial x}\right)^j = \sum_{j=0}^{\infty} A^j, \tag{30}$$

where it is assumed that

$$\lim_{t\to\infty} A^j = 0 \tag{31}$$

The limiting behavior of A in (31) is the analogue of $\beta(1+\partial g/\partial x)<1$ (cf (28)) in the case of a single resource. The limiting sum in (30) with condition (31) met is

$$[I - A]^{-1} = \left[I - \beta\left(I + \frac{\partial g}{\partial x}\right)\right]^{-1} \tag{32}$$

The same relationship exists for the depreciation function, i.e.,

$$\sum_{j=0}^{\infty} \beta^j \left(I - \frac{\partial D}{\partial y}\right)^j = \left[I - \beta\left(I - \frac{\partial D}{\partial y}\right)\right]^{-1} \tag{33}$$

Therefore, equations (26) and (27) are also applicable in our generalized framework.

Applications may often call for a model capable of handling several interrelated natural resources. Biologically interacting species of fish in a commercial fishery is an economically important area of application. Another common situation is interrelated groundwater aquifers. In fact, many single groundwater basins can be advantageously partitioned into highly interrelated sub-basins for more detailed analysis, both from the point of view of withdrawal rates and investment in pumping capacity. In many respects the number of separate resources would be arbitrarily determined vis-à-vis the number of sub-basins specified.

Appendix

In this Appendix, we take a brief look at sufficient conditions for a maximum in relation (4), consider existence of the Lagrange multipliers in the Kuhn-Tucker Theorem, and examine some problems arising under an infinite planning horizon.

We begin under the assumptions of a finite planning horizon and existence of the Lagrange multipliers in the Kuhn-Tucker Theorem, and specify a set of conditions that are sufficient for (4) to be a maximum. The assumptions that $g(u_t, x_t)$ and $h(x_t, y_t)$ are concave and that $D(u_t, v_t, y_t)$ is convex imply convexity for the constraint equations in (4-a) through (4-c). We have also assumed a concave criterion function by the assumed concavity of $B_t(u_t, v_t, x_t, y_t)$, $t=0, 1, \ldots,$ T-1, and $\psi(x_T, y_T)$.

We note that the multipliers, $\{\mu_t\}$ and $\{\lambda_t\}$, can be negative in the formulation depicted by (4) because of the equalities in constraints (4-a) and (4-b). If these equations were replaced by inequalities,

$$x_{t+1} \leq x_t + g(u_t, x_t),$$
$$y_{t+1} \leq y_t - D(u_t, v_t, y_t),$$
$$t = 0, 1, \ldots, T-1,$$

all the multipliers would be nonnegative by the necessary conditions of Kuhn and Tucker. But the implication of inequalities replacing the equalities of (4-a) and (4-b) is costless disposal of any redundant resource or capital stocks. Of course, costless disposal of surplus resource or capital stocks may be extremely restrictive in some cases. For example, surplus groundwater stocks would usually imply a drainage problem. Therefore, an additional restriction must be imposed before the Kuhn-Tucker conditions imply

sufficiency as well as necessity, even with our assumptions of concavity and convexity on the functions of (4). This restriction is that the multipliers $\{\lambda_t\}$ and $\{\mu_t\}$ be positive (see Garland Hadley, pp. 192–93), which may imply free disposal. In other words, free disposal is a sufficient, but not a necessary, condition for $\{\lambda_t\}$ and $\{\mu_t\}$ positive.

Necessary and sufficient conditions for existence of the Lagrange multipliers and thus application of the Kuhn-Tucker theorem, were proved by David Gale. Let w be a vector of variables added to the right-hand side of the constraints (4-a) through (4-c) replacing the zeros. We define $\phi(w)$ as the supremum of the criterion function in (4) for this modification in the constraints. In other words, $\phi(w)$ is the criterion function at its supremum for the family of optimization problems obtained by substitution of the vector w for the null vector in the right-hand side of the constraints (4-a) through (4-c).

Gale's main result for our purposes is that the Lagrange multipliers in (5) exist if, and only if, $\phi(w)$ has bounded steepness at $w=0$. Bounded steepness is defined as a finite value for

$$s = \sup_{w \in W}[\phi(w) - \phi(0)]/\|w\|,$$

where $\|w\|$ is some arbitrary norm of the vector w. The set W is all vectors w that are consistent with the modified constraint set on $\{u_t, v_t, x_t, y_t\}$ after replacing the null vector in (4-a) through (4-c) with w.

It would seem that violation of this condition for existence would imply a logical error in the specification of the economic model. If the steepness, s, is unbounded, discounted social benefits per unit of input (in some aggregated measure of the natural resource and capital) can be increased without limit. Therefore, it seems appropriate to assume existence of the multipliers, and should difficulty be encountered in applications, the problem is almost certainly a wrong specification of the constraint set.

The case of an infinite planning horizon raises problems of considerable difficulty. We come to grips with them in only a superficial manner by exploiting previous analyses that are applicable, viz., the dynamic programming approach of Bellman (1957). Bellman proved existence and uniqueness for our criterion function under conditions more restrictive, in some ways, then we have assumed thus far. Also, a method of successive approximations is provided by his analysis that gives uniform convergence for the criterion function. This method of successive approximations can be applied to a sequence of models identical to (4).

At this point, let us define a function $f_T(x_0, y_0)$ which is the value of the objective function in (4) at its maximum value, given that initial stocks are x_0 and y_0, viz., $f_T(x_0, y_0)$ is the solution of a family of optimization problems defined by (4). The following more restrictive assumptions are made with respect to this optimization problem:

(A-1) a) $B(u, v, x, y)$ is not explicitly a function of time (subscript t removed);
b) x and y are confined to a bounded region, say R;
c) The difference equations (4-a) and (4-b) are consistent with $(x, y) \in R$;
d) $B(u, v, x, y)$ is uniformly bounded for all values of its arguments belonging to the constraint set specified in (4) and $(x, y) \in R$.

The theorem of interest to us states that under our new assumptions, $f_T(x_0, y_0)$ converges as $T \to \infty$ to the functional equation,

$$\text{(A-2)} \quad f(x, y) = \operatorname*{Max}_{u,v}[B(u, v, x, y) + \beta f(x + g(u, x), y - D(u, v, y))],$$

which has a unique solution (Bellman 1957, p. 121). Further, the solution may be obtained as the limit of a sequence $\{f_n(x, y)\}$, defined as (A-3), shown below. Also, the sequence converges uniformly to its limit. As a result of our more restrictive assumptions (continuity and the bounded region R), we were able to dispense with the supremum in Bellman's an alysis and use a maximum.

The reader may be disturbed that we dropped the terminal function $\psi(x, y)$ by our definition of $f_0(x, y)$ in (A-3). This is of

$$\text{(A-3)}\quad f_0(x, y) = \underset{u,v}{\text{Max}}\, B(u, v, x, y)$$
$$\vdots$$
$$f_n(x, y) = \underset{u,v}{\text{Max}}\, [B(u, v, x, y) + \beta f_{n-1}(x + g(u, x), y - D(u, v, y))], \qquad n = 1, 2, \ldots.$$

no consequence, however, since the final value of resource and capital stocks are ultimately of infinitesimal importance as $T \to \infty$. We could just as well have used $f_0(x, y) = \psi(x, y)$ with suitable boundedness constraints on ψ.

With the additional restrictions in (A-1) appended to the optimization model in (4), a sequence of solutions can be used to approximate the infinite horizon solution to any desired degree of accuracy. The method is to solve the T-horizon model for *many* initial stocks (x_0, y_0), which permits an estimate of $f_T(x, y)$. The magnitude of T is arbitrary and in applications would be determined to considerable extent by computer capacity required in non-linear programming algorithms. The next step is to set $\psi(x, y) = f_T(x, y)$ in the T-period horizon model and solve the problem again for many pairs (x_0, y_0), which will yield an estimate of $f_{2T}(x, y)$. The next iteration uses $\psi(x, y) = f_{2T}(x, y)$ to generate an estimate of $f_{3T}(x, y)$ etc. Thereby, approximation to $f(x, y)$ is achieved to any degree desired by virtue of the uniform convergence of $f_n(x, y)$ to $f(x, y)$ as $n \to \infty$.

The restriction (A-1a) is not very burdensome because projections of the future become so nebulous after about 50 years that the best guess is to assume $B(u, v, x, y)$ constant for $t > M$, where M is some finite year in the future after which changes are too vague to project. After getting a sufficiently accurate estimate of $f(x, y)$ using the invarient function $B(u, v, x, y)$, set $\psi(x, y) = f(x, y)$, then solve the model exemplified by (4) with $T = M$, where time enters the social benefit function explicitly. If necessary for computational purposes, two or more steps could be used in solving this M-period problem.

In Section III we are concerned with the necessary conditions for a maximum of social benefits as the planning horizon becomes infinite. Therefore, we need to show that the necessary conditions of (19) and (20) hold for the limiting case where $T \to \infty$, which would imply an infinite number of variables in (4). This result follows almost directly from Bellman's theorem and the function, $f(x, y)$, in (A-2). We know that $f(x, y)$ exists and is unique under the additional constraints of (A-1). Let us think of the optimization problem in (4), modified by the additional constraints of (A-1), as that associated with an infinite planning horizon where the terminal value function $\psi(x_T, y_T)$ measures value for $t \geq T$. Then obviously, the terminal value function is $f(x, y)$, viz.,

$$\psi(x_T, y_T) = f(x_T, y_T).$$

Going to the necessary conditions for (4) to be a maximum (under the assumption $x_t, y_t > 0$, $t = 0, 1, \ldots$), equations (19) and (20), it is clear that we can approximate the necessary conditions under an infinite planning horizon as closely as we please by taking T arbitrarily large, as long as the terms of the summations in (19) and (20) ultimately converge to zero. The discount factor, $0 < \beta < 1$, in conjunction with boundedness assumptions, etc., imposed in (4) and (A-1) assure such convergence.

REFERENCES

R. E. Bellman, *Dynamic Programming*, Princeton 1957.

———, *Adaptive Control Processes: A Guided Tour*, Princeton 1961.

G. Brown, Jr. and C. B. McGuire, "A Socially Optimum Pricing Policy for a Public Water Agency," *Water Resources Research*, 1st quarter 1967, *3*, 33–43.

O. R. Burt, "Groundwater Management Under

Quadratic Criterion Functions," *Water Resources Research*, 3d quarter 1967, *3*, 673–82.

J. A. Crutchfield and A. Zellner, *Economic Aspects of the Pacific Halibut Industry*, Fishery Industrial Research I, U.S. Department of the Interior, Washington 1962.

R. Cummings and O. Burt, "The Economics of Production from Natural Resources: Note," *Amer. Econ. Rev.*, Dec. 1969, *59*, 985–90.

P. Davidson, "Public Policy Problems of the Domestic Crude Oil Industry," *Amer. Econ. Rev.*, Mar. 1963, *53*, 85–108.

D. Gale, "On Optimal Development in a Multisector Economy," *Rev. Econ. Stud.*, Jan. 1967, *34*, 1–18.

H. S. Gordon, "The Economic Theory of a Common Property Resource: The Fishery," *J. Polit. Econ.*, Apr. 1954, *62*, 124–42.

R. L. Gordon, "A Reinterpretation of the Pure Theory of Exhaustion," *J. Polit. Econ.*, June 1967, *75*, 274–86.

L. C. Gray, "Rent Under the Assumption of Exhaustibility," *Quart. J. Econ.*, May 1914, *28*, 466–89.

G. Hadley, *Nonlinear and Dynamic Programming*, Reading 1964.

H. Hotelling, "The Economics of Exhaustible Resources," *J. Polit. Econ.*, Apr. 1931, *39*, 137–75.

J. V. Krutilla, "Conservation Reconsidered," *Amer. Econ. Rev.*, Sept. 1967, *57*, 777–86.

H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium On Mathematical Statistics and Probability*, Berkeley 1951, pp. 481–92.

J. Quirk and V. Smith, "Dynamic Economic Models of Fishing," in A. Scott, ed., *Proceedings of the H. R. MacMillan Fisheries Economics Symposium*, Vancouver 1969.

A. Scott, "The Fishery: The Objectives of Sole Ownership," *J. Polit. Econ.*, Apr. 1955, *63*, 116–24.

———, "The Theory of the Mine under Conditions of Certainty," in M. Gaffney, ed., *Extraction Resources and Taxation*, Madison 1967.

V. Smith, "Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.

R. Turvey, "Optimization in Fishery Regulation," *Amer. Econ. Rev.*, Mar. 1964, *54*, 64–76.

The Effect of Interest Rates on Aggregate Consumption

By WARREN E. WEBER*

According to J. M. Keynes, . . . "The usual type of short-period fluctuation in the rate of interest is not likely . . . to have much *direct* influence on spending either way" (p. 93). The classical view that consumption is a function of the rate of interest has received relatively little attention in empirical studies of the consumption function.[1] Given the importance of aggregate demand in the economy we should further examine the question whether monetary policy can directly affect aggregate consumption through the rate of interest.

This study examines the relationship between consumption and the rate of interest. The specific aggregate consumption functions tested are derived under the assumption that individuals maximize utility over a multiperiod horizon. In this respect, the approach is similar to that of Franco Modigliani and Richard Brumberg (1953, 1954) and Milton Friedman, except that interest rate effects are not assumed away. The empirical test of whether interest rates are a significant determinant of aggregate consumption is performed with aggregate data of the United States for the period 1930–65, omitting 1941–46. We find that interest rates do significantly affect aggregate consumption and that an increase in the rate of interest increases aggregate consumption, other things being equal.

The analysis is divided into six sections. In the first we develop an equation for the consumption of an individual. This equation includes the rate of interest as an independent variable. In the second we examine the role of expectations in the model, and in the third we develop an aggregate consumption function. In the fourth section we discuss some statistical problems which arise in the analysis. Then we empirically test the theory in the fifth section and conclude by discussing some of the implications of our analysis.

I. *The Representative Consumer*

Suppose that a representative consumer in the economy begins his economic life at age 0, lives L years, and desires to leave no bequests. When he is at age i, he has a remaining lifetime of L-i-1 years over which to allocate his consumption. In addition, assume that he has a utility function which is a monotonic transformation of the CES function.[2] Specifically, assume

* Assistant professor of economics, Virginia Polytechnic Institute. I am indebted to Michael Lovell, Robert Lucas, Morris De Groot, Vittorio Bonomo, and an anonymous referee for criticism and helpful suggestions. Any errors which remain are mine.

[1] Two recent exceptions are the studies by Arnold Zellner et al. and by C. Wright. Laurits Christensen examines the effect of the rate of return on capital on consumption, and William Lott examines the effect of interest rates on the average propensity to consume out of permanent income.

[2] The technique of assuming that the utility function of the representative consumer has a specific form has been used before. In his analysis of commodity demand, Richard Stone (1954) derives a "linear expenditure system" using the utility function given by Robert Geary. The linear expenditure system is also used by Stone (1965, 1966); Stone, A. Brown, and D. A. Rowe; and Arthur Goldberger and T. Gamaletsos. "Direct addilog" and "indirect addilog" utility functions are proposed and used by Hendrik Houthakker in his analysis of British demand. Previously, Christensen assumes an intertemporal CES utility function for the consumer and a specific production function for the firm in his

that his utility function is

$$(1)\quad u(c_i, \ldots, c_{L-1}) = \alpha \sum_{j=i}^{L-1} (1+\gamma)^{j-i} c_j^{-\beta},$$

where c_j is his total consumption at age j and α, β, and γ are parameters.[3] We assume that $\alpha\beta<0$ and that $-1<\beta<\infty$ to assure that the marginal utility of consumption is positive and that the utility function is quasi-concave.[4] We also restrict γ to be greater than -1. In (1) we can identify $\sigma=1/(1+\beta)$ as the partial intertemporal elasticity of substitution of consumption and γ as the consumer's subjective rate of discount for future consumption (his rate of time preference). Normally, $\gamma<0$.

We now let y_j denote the representative consumer's (after-tax) labor income at age j, w_i his non-human wealth at the beginning of age i, and r the rate of interest at which he can borrow and lend. Using this notation his budget constraint is

$$(2)\quad \sum_{j=i}^{L-1} \left(\frac{1}{1+r}\right)^{j-i} c_j = v_i,$$

where

$$v_i = w_i + \sum_{j=i}^{L-1} \left(\frac{1}{1+r}\right)^{j-i} y_j$$

The individual's discounted consumption must be equal to his total human and non-human wealth at the beginning of age i.

Maximizing (1) subject to the wealth constraint (2) yields the following necessary conditions:

$$(3)\quad \begin{aligned} -\alpha\beta(1+\gamma)^{j-i} c_j^{-\beta-1} + \lambda\left(\frac{1}{1+r}\right)^{j-i} &= 0 \\ j = i, \ldots, L-1 \\ \sum_{j=i}^{L-1}\left(\frac{1}{1+r}\right)^{j-i} c_j &= v_i \end{aligned}$$

where λ is a Lagrangian multiplier. Solving (3) yields the L-i equations which express the individual's lifetime consumption plan as formulated at age i. In particular, his demand for current consumption is

$$(4)\quad c_i = g(r, i)\, v_i$$

where we let

$$g(r, i) = \left[\sum_{j=i}^{L-1} [(1+\gamma)(1+r)^{-\beta}]^{\sigma(j-i)}\right]^{-1}$$

The individual's demand for current consumption is proportional to his total wealth. His marginal propensity to consume out of total wealth $g(r, i)$ depends upon his partial elasticity of substitution, his rate of time preference, and the rate of interest. Consequently, the interest rate affects both the individual's total wealth and his marginal propensity to consume out of it.[5]

II. *The Role of Expectations*

The theory of the representative consumer developed above cannot be tested directly because we cannot observe future labor income. In order to test our theory, we adopt an approach to unobservable variables often used in the econometric

---

analysis of the rate of return on capital and total private saving.

[3] The individual's consumption is defined to be his purchases of nondurable goods and services plus the services of durables which he used during the period. Thus, we follow J. R. Hicks, Modigliani and Brumberg (1953, 1954), and Friedman in recognizing that purchases of durables are a capital expenditure and not necessarily equal to the consumption of durables.

[4] See Kenneth Arrow and Alain Enthoven.

[5] We can obtain Modigliani and Brumberg's equation for current consumption as a special case of (4). They assume that the "... hypothetical prototype plans to consume his income at an even rate throughout the balance of his life" (1954, p. 397). Thus, his marginal propensity to consume out of total wealth is independent of the rate of interest and equal to $1/(L\text{-}i)$. In addition, they assume that the rate of interest is zero and constant, so that the consumer's total wealth is unaffected by the rate of interest. Consequently their result is a special case of (4) which prevails if $r=\gamma=0$.

literature. We will include specific assumptions relating the unobservable variables to observable ones as part of the maintained hypothesis.[6]

We assume that the consumer anticipates that future labor income will grow or decline at some constant rate ξ over the remainder of his lifetime. Thus, his anticipated stream of future labor income is given by

$$(5) \quad y_j = (1 + \xi)^{j-i} y_i \qquad \xi > -1, \quad j = i + 1, \ldots, L - 1$$

The expected rate of growth of labor income ξ will be estimated in the course of the empirical analysis.[7]

We now face the question of what observable market interest rate or rates should be used for r in our empirical analysis. The consumer faces a number of interest rates in the market. In formulating his consumption plan he should take these various rates as well as his anticipations concerning their future levels into account.

In light of this fact, r should be considered an average effective interest rate or as an index of the interest rates which the consumer faces. This average effective interest rate may not be closely approximated by any current market rate. Thus, we cannot simply substitute a particular market rate for r. Instead, we will allow the consumer's average effective rate to be determined as a part of the empirical analysis. We will assume that the representative consumer forms an "estimate" of the average effective interest rate he faces according to

$$(6) \qquad r = \theta_1 + \theta_2 r^m,$$

where r^m is a particular market rate of interest, and θ_1 and θ_2 are parameters to be estimated in the empirical analysis.

The consumer constructs his interest rate index on the basis of a particular market interest rate, but he adjusts this rate using θ_1 and θ_2 to indicate the pattern of rates he faces in the market. The formulation (6) allows the consumer's average effective interest rate to have a larger or smaller mean and variance (larger or smaller level and degree of fluctuation) than the market rate chosen as the basis for the estimate. Thus, there is a major difference between using (6) and using a particular market interest rate for r in the empirical analysis.

In (6) the parameter θ_1 is a "risk premium" which the consumer has to pay in order to borrow or a "floor" under the rate of return which he can earn on his assets. This parameter may be greater than zero because of the existence of highly liquid, relatively riskless interest-bearing assets such as savings accounts. Also, uncertainty about the consumer's future income stream or future price levels may tend to make the parameter positive. The parameter θ_2 measures the sensitivity of the consumer's rate of interest to changes in the market rate. It will be equal to zero if the consumer's expected rate of interest is insensitive to changes in the market rate. It will be less than one if the consumer does not fully adjust his interest rate expectations to changes in the market rate. One reason that the consumer may not fully adjust his expectations is that he may have entered into long-term contracts at interest rates which differ from the current market rate. The parameter θ_2 may also pick up the effect of tax rates upon the rate of return that an individual may earn on his nonhuman wealth.[8] Such a tax effect could also

[6] In previous studies of consumption this approach is used by Watts; Ando and Modigliani; Arena; and Zellner, Huang, and Chau.

[7] The effect on the model of replacing (5) with an adaptive expectations hypothesis about the way in which the consumer forms his expectations is discussed in my doctoral thesis.

[8] I am indebted to a referee for pointing out that tax rates might effect the individual's average effective rate of interest.

cause θ_2 to be less than one. Of course, since θ_2 is a constant, it will not pick up the effects of any changes in tax rates which may have occurred over time.[9]

III. *An Aggregate Consumption Function*

By including the assumptions about interest rate and future labor income expectations in the model, we can express the consumer's demand for current consumption in terms of only observable variables. Specifically, substituting (5) and (6) into (4) we obtain

$$(7)\quad c_i = g(r^m, i)\cdot\left[w_i + y_i \sum_{j=i}^{L-1}\left(\frac{1+\xi}{1+\theta_1+\theta_2 r^m}\right)^{j-i}\right]$$

where

$$g(r^m, i) = \left[\sum_{j=i}^{L-1}\left[(1+\gamma)(1+\theta_1+\theta_2 r^m)^{-\beta}\right]^{\sigma(j-i)}\right]^{-1}$$

We find that the consumer's demand for current consumption is still proportional to his total wealth. However, his total wealth now depends on his beginning of period non-human wealth, his current labor income, and the market rate of interest as well as the parameters θ_1, θ_2, and ξ. His marginal propensity to consume now depends upon the market rate of interest as well as the parameters γ, θ_1, θ_2, and β. The individual's demand for current consumption will be unaffected by changes in the market rate of interest if $\theta_2=0$.

Now suppose that all consumers behave like a representative consumer at economic age e. Then, letting capital letters denote the aggregate values of the variables and adding a subscript to denote the time period, we know that aggregate consumption at time t is

$$(8)\quad C_t = g(r_t^m, e)\cdot\left[W_t + Y_t \sum_{j=0}^{L-e-1}\left(\frac{1+\xi}{1+\theta_1+\theta_2 r_t^m}\right)^{j}\right]$$

We note that this aggregate consumption function could also have been obtained by assuming that (7) held in the aggregate.[10]

IV. *A Statistical Digression*

Before performing the empirical estimation we must resolve an econometric problem which arises in connection with the empirical estimation. This problem arises from the fact that (8) is not identified in the sense that it is not possible to estimate two or more of the parameters because their effects cannot be separated empirically. In other words, it is possible to express the model in terms of fewer than the original number of parameters.

The fact that (8) is not identified is easily seen if we define three constants k_1, k_2, and k_3 such that

$$(9)\quad \begin{aligned} k_1 &= (1+\gamma)^{-1/\beta}(1+\theta_1) \\ k_2 &= (1+\gamma)^{-1/\beta}\theta_2 \\ k_3 &= (1+\theta_1)(1+\xi)^{-1} \end{aligned}$$

These three equations can be solved to obtain

$$\theta_2(1+\xi)^{-1} = k_2 k_3 k_1^{-1}$$

Consequently, using (9) we can eliminate γ, θ_1, θ_2, and ξ from (8) and express it in terms of β, k_1, k_2, and k_3. Specifically we

[9] The parameters θ_1 and θ_2 can be interpreted in another way. Suppose we rewrite (6) as

$$r = \eta + \theta_2(r^m - \eta),$$

where $\eta=\theta_1(1-\theta_2)^{-1}$. Thus, η can be viewed as the consumer's long-run or "normal" interest rate. It is unaffected by currently prevailing market conditions. The parameter θ_2 measures how much the consumer adjusts his current expected interest rate to differences between the market rate and his normal rate. His current expected rate always equals η if $\theta_2=0$. I am indebted to Vittorio Bonomo for this interpretation.

[10] The effect on the model of using the aggregation assumption that total wealth is equally distributed among age groups, rather than the assumption that all consumers are the same economic age, is discussed in my thesis.

TABLE 1

Parameter	Durand's Basic Yields Years to Maturity			Moody's Baa	Average Annual Yield Time and Savings Deposits In Com. Banks
	1	5	30		
β	6.288	2.599	1.427	6.675	2.778
σ	0.1372	0.2779	0.4120	0.1303	0.2647
k_1	1.083	1.089	1.083	1.016	1.079
k_2	0.1953	0.3737	0.9819	0.5904	0.4543
k_3	1.145	1.127	1.093	1.037	1.109
Sum of Sq. Residuals	322.7	289.8	291.0	222.3	235.5
Test Statistic (Ω)	11.5	14.8	14.6	22.7	21.0

obtain

$$C_t = g^*(\overset{m}{r}_t, e)\left\{W_t + Y_t \sum_{j=0}^{L-e-1} \left[\frac{k_1}{k_3(k_1 + k_2\overset{m}{r}_t)}\right]^j\right\} + \epsilon_{1t}, \tag{10}$$

where

$$g^*(\overset{m}{r}_t, e) = \left[\sum_{j=0}^{L-e-1} (k_1 + k_2\overset{m}{r}_t)^{-\beta\sigma j}\right]^{-1}$$

and ϵ_{1t} is a normally distributed error term with mean 0 and variance ν_1^2.

The estimation of (10) yields estimates of the five parameters β, k_1, k_2, k_3 and ν_1^2. However, we cannot substitute these estimates into (9) and solve to obtain indirect estimates of the four original parameters γ, θ_1, θ_2, and ξ. Nevertheless, we can still obtain some information about these parameters. For example, $\theta_1\{\gtreqless\}\xi$ as $k_3\{\gtreqless\}1$. And, more importantly, we can still test for the significance of interest rates as a determinant of aggregate consumption. Because of the restrictions that $\gamma > -1$ and $\xi > -1$, k_2 will be equal to zero only when $\theta_2 = 0$. Thus, to test for the significance of interest rates we test the null hypothesis that $k_2 = 0$ against the alternative hypothesis that $k_2 > 0$.

All hypotheses will be tested using the likelihood ratio criterion. The test statistics are calculated in the following manner. When the error terms are normally distributed, we know that the likelihood ratio (Λ) is

$$\Lambda = (SS_a/SS_n)^{T/2}, \tag{11}$$

where T is the number of observations and SS_a and SS_n are the minimum sum of the squared residuals under the alternative and null hypotheses, respectively. Asymptotically, $-2\ log_e\ \Lambda$ is distributed as chi-square with degrees of freedom equal to the number of restrictions being tested. Under the assumption that our sample is sufficiently large for the asymptotic results to be a valid small sample approximation, we can test a particular null hypothesis at a specified level of significance by comparing the value of the test statistic

$$\Omega = -T(log_e SS_a - log_e SS_n) \tag{12}$$

against the percentiles of the chi-square distribution with degrees of freedom equal to the number of restrictions being tested.

V. *Empirical Results*

The results of the empirical analysis are presented in Table 1 above. It is performed with five different interest rate series.[11]

[11] The first three are basic yields on corporate bonds

More than one interest rate series is used to determine if the choice of interest rate series affects the results in any important respect. The data series used in the analysis will be provided on request by the author. Details on the construction of the aggregate non-human wealth, labor income, and consumption series can be found in my doctoral thesis. All variables except the interest rate are expressed in real terms by dividing by the Consumer Price Index.[12]

In the bottom row of Table 1, we present the test statistics for a likelihood ratio test of the null hypothesis that $k_2=0$ against the alternative hypothesis that $k_2>0$.[13] All of the resulting values of Ω are greater than 10.8, the 0.999 percentile of the chi-square distribution with one degree of freedom. Therefore, we reject the null hypothesis that interest rates do not affect consumption at the 0.001 level of significance for all interest rate series.[14]

Even though the choice of an interest rate series does not affect the above hypothesis tests, there are differences in the explanatory power of the various interest rate series. Suppose the better the explanatory ability of an interest rate, the more reliable consumers consider it as an indicator of the interest rates they face. In this case, our analysis shows that they consider the yield on time and savings deposits and on Baa corporate bonds to be the best indicators of conditions in their lending and borrowing markets. It also indicates that they consider corporate bonds yields to be more reliable indicators the more risky the bond.[15]

1. *Economic Significance of the Parameter Estimates*

We are only able to obtain an estimate of one of the five original parameters in the model; namely, β. The estimates of β are uniformly greater than one, so that our estimates of the partial elasticity of substitution $[\sigma=1/(1+\beta)]$ are uniformly less than 0.5. The partial elasticity of substitution indicates the ease with which consumption in one time period can be sub-

with one, five, and thirty years to maturity obtained from David Durand and *The Economic Almanac*. The fourth series used is Moody's Baa corporate bond rate obtained from *Banking and Monetary Statistics*; the *Supplement to Banking and Monetary Statistics, Section 12*; and various issues of the *Federal Reserve Bulletin*. The fifth series is the average annual yield on time and savings deposits in commercial banks obtained from the *Savings and Loan Fact Book*.

[12] The non-linear estimations are carried out using a direct-search routine described by R. Hooke and T. Jeeves. In the estimations the planning horizon of consumers (L-e) is set equal to thirty years. This is approximately equal to the life expectancy of an individual whose age is the median age of household heads over the observation period. The median age of heads of households was approximately 45 between 1930 and 1965. Between 1950 and 1965 the life expectancy of a 45 year old varied between 28.8 and 29.7 years. In 1930 and 1940, the life expectancy of a 40 year old was 29.22 years and 30.03 years, respectively. The sources of these data are *Historical Statistics of the United States* and various issues of the *Statistical Abstract of the United States*.

[13] These test statistics are calculated in the following manner. For each interest rate series the minimum sum of the squared residuals under the null hypothesis (SS_n) is obtained by substituting $k_2=0$ into (10) and then estimating the resulting consumption function

$$C_t = a_1W_t + a_2Y_t + e_{2t},$$

where $a_1=g^*(0, e)$, $a_2=a_1\sum_{j=0}^{L-e-1} k_3^{-j}$, and e_{2t} is a normally distributed error term with mean 0 and variance v_2^2. Estimating this equation as a linear regression we find

$$C_t = 0.07743W_t + 0.6117Y_t + e_t.$$

$SS_n=473.9$. This result is very similar to that obtained by Ando and Modigliani for the period 1929–59, also omitting 1941–46. We then substitute the values of SS_a and SS_n into (12).

[14] The chi-square distribution with one degree of freedom is used because the null hypothesis specified only one restriction on the parameters. However, under the alternative we estimate five parameters, whereas under the null hypothesis we estimate only three parameters. This suggests that two degrees of freedom may be appropriate or that the limiting chi-square distribution may be inappropriate since $k_2=0$ is a boundary point of the parameter space. See Weber (p. 108) for a justification of the use of the chi-square distribution with one degree of freedom.

[15] To check this conclusion we also performed the analysis with the yield on Moody's Aaa corporate bonds. We found that it fit better than the Durand's yields and worse than the Baa yield.

stituted for consumption in another time period to maintain the same level of utility. Since the degree of substitutability increases with σ, our results indicate that consumers view present and future consumption to be poor substitutes.

We cannot obtain direct estimates of θ_1, θ_2, γ, and ξ. However, suppose we accept the interpretation that θ_1 is an interest rate floor or a risk premium. Then we have some basis for assigning an a priori value to this parameter, and we can substitute our estimates of β, k_1, k_2, and k_3 into (9) and solve for the other three original parameters. We let $\theta_1 = 0.02$ which is approximately the mean of the average annual yield on time and savings deposits in commercial banks over the observation period. Substituting into (9) we obtain the values of γ, ξ, and θ_2 given in Table 2 below.

The estimates of the *time preference parameter* (γ) range from -0.3140 to 0.0162 indicating that consumers discount future consumption at rates between -31.4 percent and 1.5 percent. This evidence indicates that consumers are impatient (that is, they prefer present consumption to future consumption) even though we obtain negative time preference with the Moody's Baa rate.

The estimates of the *anticipated rate of labor income growth* (ξ) range from -0.1092 to -0.0162 indicating that consumers expect their labor income to decline at an annual rate of between 10.9 percent to 1.6 percent per year. Thus, if interest rate effects are neglected, consumers' human wealth is approximately 9 to 24 times their annual labor income. These estimates are somewhat larger than William Hamburger's finding that " . . . the average wage earner's anticipations of labor-based consuming power is about six times as high as his annual wage" (p. 10).

However, the secular growth of labor income would lead us to expect the estimates, of ξ to be positive. Our estimates of course, are influenced by the length of consumers' planning horizons. If the median age of heads of households is 45 and they have a life expectancy of thirty years, then such a horizon may be long enough to include retirement. Since retirement can mean a substantial fall in labor income, its effect is to reduce the estimates of ξ.

To check this explanation we reestimate (10) under the assumption that the last ten years of consumers' planning horizons are retirement years in which no labor income is received. Using Moody's Baa, since it provides the best fit in Table 1, we obtain the following result

Parameter	Estimate
β	2.348
k_1	1.016
k_2	0.7797
k_3	1.003
Sum of squared residuals	221.9

Evaluating these results in the same manner as those in Table 1, we find that $\xi = 0.0096$, so that in this case consumers do expect that labor income will grow over

TABLE 2

Parameter	Durand's 1	Durand's 5	Durand's 30	Moody's Baa	Time and Savings Deposits Yield
γ	-0.3140	-0.1564	-0.0820	0.0274	-0.1446
ξ	-0.1092	-0.0949	-0.0668	-0.0162	-0.0803
θ_2	0.1839	0.3500	0.9248	0.5928	0.4295

the remaining twenty years of their working lifetime.[16]

The estimates of *the sensitivity of consumers' average effective interest rate to changes in the market rate* (θ_2) are uniformly less than one indicating that consumers do not fully adjust their interest rate index to changes in the market rate. However, they adjust more quickly the longer the maturity of the bond. They also adjust more quickly to changes in the Baa rate than to changes in the average annual yield on time and savings deposits. The θ_2 also determine the variance of consumers' average effective interest rate from the market rates upon which these estimates are based. The variance of the Durand's basic yields decreases as the maturity increases. However, the variance of the average effective interest rate based on the Durand's yields increases as the maturity of the Durand's yield increases. The average effective interest rate with the highest variance is that based on Moody's Baa, which has the highest variance and provides the best fit of the five interest rate series.

2. *Tests of Two Hypotheses*

We now present the results of two hypothesis tests. The first is whether the partial intertemporal elasticity of substitution is equal to one, and the second is whether consumers' average effective rate of interest is equal to the currently prevailing market rate.

For the first test we can obtain the minimum sum of the squared residuals under the alternative hypothesis from Table 1. However, we cannot obtain the minimum sum of the squared residuals under the null hypothesis by substituting $\beta=0$ into (10) and then estimating the resulting equation because we had to assume that $\beta\neq 0$ to assure that (1) was quasi-concave.

Now we know that if a representative consumer has a utility function of the form

$$(13)\qquad u(c_i, \ldots, c_{L-1}) = \sum_{j=i}^{L-1} \alpha_j \log c_j,$$

where the α_j are parameters, the partial elasticity of substitution is equal to one.[17] Maximizing this utility function subject to the wealth constraint (2), substituting (5) and (6) into the resulting demand for current consumption equation, and aggregating as above, we find that aggregate consumption at time t is

$$(14)\qquad C_t = \alpha_* \left[W_t + Y_t \sum_{j=0}^{L-*-1} \cdot \left(\frac{1}{k_3 + k_4 r_t^m} \right)^j \right] + \epsilon_{3t},$$

where $k_4=\theta_2(1+\xi)^{-1}$, and ϵ_{3t} is a normally distributed error term with mean 0 and variance ν_3^2. When the partial elasticity of substitution is equal to one, the marginal propensity to consume out of total wealth no longer depends upon the rate of interest even though total wealth still depends upon the rate of interest.

Consequently, we find the minimum sum of the squared residuals under the null hypothesis $\beta=0$ by estimating (14). Using Moody's Baa since it provides the best fit, we find that $SS_n=473.9$.[18] Since $SS_a=$

[16] Equation (10) was also reestimated shortening the entire planning horizon to fifteen and then again to five years. In both cases ξ was positive. However, we also obtained extremely large γ, indicating that consumers are highly negatively impatient and strongly prefer future consumption almost to the total exclusion of present consumption. This result indicates that part of the relevent planning horizon for consumers was omitted in the analysis. Thus, this evidence supports the explanation that a negative ξ is obtained due to the fact that consumers' working horizons are shorter than their planning horizons and the use of a thirty year planning horizon for consumption.

[17] We note that the utility function (13) is a monotonic transformation of the Cobb-Douglas production function. The Cobb-Douglas is the limiting form of *CES* function when $\sigma=1$.

[18] We find that the sum of the squared residuals is minimized when $k_4=0$. Thus, the results obtained by estimating (14) are identical to those given in fn. 14 above.

$$(15)\quad \frac{\partial C_t}{\partial r_t^m} = k_2\beta\sigma g^*(r_t^m, e)^2\left[\sum_{j=0}^{L-e-1} j(k_1 + k_2 r_t^m)^{-\beta\sigma(j-1)}\right]\cdot\left\{W_t + Y_t\sum_{j=0}^{L-e-1}\left[\frac{k_1}{k_3(k_1+k_2r_t^m)}\right]^j\right\} - g^*(r_t^m, e)Y_t\sum_{j=0}^{L-e-1} j\left[\frac{k_1}{k_3(k_1 + k_2r_t^m)}\right]^{j-1}$$

224.9, the value of the test statistic Ω is 22.4. The 0.999 percentile of the chi-square distribution with one degree of freedom is 10.8. We reject the null hypothesis that the partial elasticity of substitution is equal to one and that interest rates do not affect the marginal propensity to consume out of total wealth.

The second hypothesis is that the interest rate which consumers expect to prevail in all future periods is equal to the currently prevailing market interest rate; i.e., that $r_t = r_t^m$. These two rates will be equal if $\theta_1 = 0$ and $\theta_2 = 1$ in (6). Substituting $\theta_1 = 0$ and $\theta_2 = 1$ into (9) we find that when the two interest rates are equal, $k_1 = k_2$, which becomes the null hypothesis for this test. The alternative hypothesis is $k_1 \neq k_2$. We find the minimum sum of the squared residuals under the null hypothesis by estimating the consumption function which results when $k_1 = k_2$ is substituted into (10). Using Moody's Baa, we obtain

Parameter	Estimate
β	3.087
$k_1 = k_2$	1.000
k_3	1.104

The minimum sum of the squared residuals is 243.6. The minimum sum of the squared residuals under the alternative hypothesis is 222.3 as given in the fourth column of Table 1. Since the difference in the sum of the squared residuals under the two hypotheses is small, we cannot reject the null hypothesis that the interest rate which consumers expect to prevail in all future periods is the market interest rate. This result indicates that for purposes of empirical analysis we were justified in using a simple relationship, such as (6), relating consumers' average effective interest rate and a particular market rate.

VI. *Implications*

We have established that given our assumptions, interest rates are an important determinant of aggregate consumption. However, we have said nothing about whether our results indicate that an increase in the interest rate will increase or decrease consumption.

We begin by taking the partial derivative of (10) with respect to r_t^m. We obtain equation (15) above. The first term on the right-hand side of (15) has the same sign as β and the second term is positive. Since our estimates of β are uniformly positive, the partial derivative is the difference of two positive numbers and its sign is indeterminate.

To obtain information on the sign of the partial derivative (15), we numerically evaluate it using the parameter estimates in Table 1. We use market rates of interest of 1 to 7 percent at intervals of 0.1 percent, values of labor income of \$200 billion to \$400 billion at intervals of \$25 billion, and values of non-human wealth of \$2 trillion to \$4 trillion at intervals of \$0.25 trillion. In all cases the derivative is positive indicating that an increase in the rate of interest leads to an increase in present consumption, other things being equal. The value of the derivative increases with interest rate, labor income, and non-human wealth.

Thus, we find that for the consumption function (10) the income effect of a change in the rate of interest is greater than the

substitution effect. When the rate of interest increases, consumers have an opportunity to maintain the same level of consumption in the future with less saving today. Consequently, they increase current consumption in response to the interest rate increase.

REFERENCES

A. Ando and F. Modigliani, "The 'Life-Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55–84.

J. J. Arena, "The Wealth Effect and Consumption: A Statistical Inquiry," *Yale Econ. Essays*, No. 2, 1963, *3*, 251–303.

K. J. Arrow and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica*, Oct. 1961, *29*, 779–800.

L. Christensen, "Saving and the Rate of Return," unpublished doctoral dissertation, Univ. California, Berkeley 1968.

D. Durand, *Basic Yields of Corporate Bonds, 1900–1942*, Nat. Bur. Econ. Res. Tech. Paper 3, New York 1942.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

R. C. Geary, "A Note on 'A Constant-Utility Index of the Cost of Living'," *Rev. Econ. Stud.*, No. 1, 1950, *18*, 65-66.

A. S. Goldberger and T. Gamaletsos, *A Cross-Country Comparison of Consumer Expenditure Patterns*, mimeo. Univ. Wis. 1967.

W. Hamburger, "The Relation of Consumption to Wealth and the Wage Rate," *Econometrica*, Jan. 1955, *23*, 1–17.

J. R. Hicks, *Value and Capital*, Oxford 1939.

R. Hooke and T. A. Jeeves, " 'Direct Search' Solution of Numerical and Statistical Problems," *Association for Computing Machinery*, Apr. 1961, *8*, 212–29.

H. S. Houthakker, "Additive Preferences," *Econometrica*, Jan. 1960, *27*, 244–57.

J. M. Keynes, *The General Theory of Employment, Interest, and Money*, London 1936.

W. F. Lott, "The Effect of Demographic Characteristics and the Interest Rate on the Consumption Function of an Economy Over Time," unpublished doctoral dissertation, N.C. State Univ. 1969.

F. Modigliani and R. Brumberg, *Utility Analysis and Aggregate Consumption Functions: An Attempt at Integration*, mimeo. 1953.

——— and ———, "Utility Analysis and the Consumption Function: An Interpretation of Cross Section Data," in K. Kurihara ed., *Post-Keynesian Economics*, New Brunswick 1954.

R. Stone, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Econ. J.*, Sept. 1954, *64*, 511–27.

———, "Models for Demand Projections," in C. R. Rao, ed., *Essays on Econometrics and Planning*, Oxford 1965.

———, *Mathematics in the Social Sciences and Other Essays*, London 1966.

———, A. Brown, and D. A. Rowe, "Demand Analysis and Projections for Britain: 1900–1970," in J. Sandee, ed., *Europe's Future Consumption*, Amsterdam 1964.

H. Watts, "Long-Run Expectations and Consumer Saving," in Dernberg, Rosette, and Watts, eds., *Studies in Household Economic Behavior*, New Haven 1958.

W. E. Weber, "The Role of Interest Rates in the Aggregate Consumption Function," unpublished doctoral dissertation, Carnegie-Mellon Univ. 1969.

C. Wright, "Some Evidence on the Interest Elasticity of Consumption," *Amer. Econ. Rev.*, Sept. 1967, *57*, 850–55.

A. Zellner, D. S. Huang, and L. C. Chau, "Further Analysis of the Short-Run Consumption Function with Emphasis on the Role of Liquid Assets," *Econometrica*, July 1965, *33*, 571–81.

Board of Governors of the Federal Reserve System, *Banking and Monetary Statistics*, Washington 1943.

———, *Fed. Res. Bull.*, Washington.

———, "Money Rates and Securities Markets," *Supplement to Banking and Monetary Statistics, Section 12*, Washington 1966.

National Industrial Conference Board, *The Economic Almanac*, New York.

U.S. Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1957*, Washington 1960.

———, *Statistical Abstract of the United States*, Washington.

U.S. Savings and Loan League, *Savings and Loan Fact Book 69*, 1969.

Expectations and the Structure of Share Prices

By BURTON G. MALKIEL AND JOHN G. CRAGG*

This paper presents the results of an empirical study of year-end common-stock prices from 1961 through 1965. The ratios of market prices to earnings are related to such factors as earnings growth, dividend payout, and various proxy variables designed to measure the risk or quality of the returns stream.

Several previous empirical studies[1] have tried to explain share prices on the basis of such variables, but these investigations were forced to rely on published accounting data and untested hypotheses about the formation of expectations. V. Whitbeck and M. Kisor were able to increase the explanatory ability of their regression by substituting the estimates of security analysts of one firm for fabricated expectations variables based on simple extrapolations of past performance. Our study tries to determine whether the goodness of fit can be improved still further by substituting the estimates from several securities firms for the expectations of a single predictor and by using a wider variety of such expectational variables. The most important of the expectational variables employed are forecasts of short-term and long-term earnings growth, estimates of the "normal earning power" of each company, and estimates of the "instability" of the earnings stream. The data used are described in Section II.

It is found in Section III that an extermely close fit to the empirical structures of share prices is obtained with the use of such expectations data. These results are also contrasted with those obtained when only historic data are used. Section III then examines further the stability and predictive power of the model over time. Section IV discusses the usefulness of the model for security selection.

I. *Specification of a Valuation Model*

In the typical valuation model, the price of a share is taken to be the present value of the returns expected therefrom. In the simplest model, the price is the sum of the present values of a stream of dividends that is assumed to grow at a constant rate, g, over time. See, for example, J. B. Williams for one of the earliest statements of the problem and M. J. Gordon for a more recent treatment. Letting P stand for the (ex dividend) price of a share, D the (annual) dividend per share in the year just past, and r the appropriate rate of discount, we have

$$P = \sum_{i=1}^{\infty} D \frac{(1+g)^i}{(1+r)^i}, \tag{1}$$

provided $g<r$. Dividing both sides of (1) by earnings per share, E, and summing the

* Princeton University and University of British Columbia, respectively. The authors are deeply grateful to John Bossons and a referee for helpful comments and to William Shaffer and James G. Mann for assistance in programming and in carrying out the computations. Thanks are also due to Raymond Hartman, Dennis Line, and Robert Lem for their invaluable help in collecting and processing the expectations data. This research was supported by the Institute for Quantitative Research in Finance and the National Science Foundation.

[1] Cross-sectional empirical studies have been undertaken by F. D. Arditti, H. Benishay, R. S. Bower and D. H. Bower, G. R. Fisher, I. Friend and M. Puckett, M. J. Gordon, F. C. Jen, M. Kisor, Jr. and A. Feuerstein, Kisor and S. Levine, and R. Ortner.

progression we obtain an expression for the price-earnings multiple

$$\frac{P}{E} = \frac{D}{E}\frac{(1+g)}{(r-g)} \tag{2}$$

The price-earnings ratio is seen to depend on the dividend payout ratio and the expected long-term growth rate of the dividend stream.

The specific model of security price-earnings ratios presented in equations (1) and (2) has several drawbacks. It is inapplicable in cases where no dividends are currently paid, it leads to an infinite value for the shares when $g>r$, and it requires projecting growth rates from now till Kingdom come.[2] Such difficulties have led several writers to formulate a finite-horizon model of share prices. See, for example, Charles Holt and Malkiel. P. F. Wendt presents a useful survey of a number of alternative models. The basic idea of the finite-horizon approach is that both dividends and earnings are assumed to grow at some rate g for N periods,[3] and then grow at a normal rate such as the growth rate for economy as a whole. This approach can be illustrated by the following very simple model.[4]

$$\frac{P_0}{E_0} = \sum_{i=1}^{N}\frac{D_0}{E_0}\frac{(1+g)^i}{(1+r)^i} + (m_s)_0\frac{(1+g)^N}{(1+r)^N}, \tag{3}$$

where $(m_s)_0$ is the average current price-earnings ratio for the market as a whole. The model in (3) appears to be highly nonlinear in the growth rate and payout ratio. Fortunately, however, a linear approximation to the true expression seems to work reasonably well for N as small as five, the period for which we have growth-rate estimates.[5]

The preceding model has abstracted entirely from the existence of risk. There are several possible ways in which risk can be represented in a valuation model. The theoretical justification for the alternatives rests on the assumptions employed.

A common way in which risk is introduced into empirical valuation models is to incorporate a term representing the (expected) variance of the future returns stream from each security. Such a procedure has been justified in two ways. First, it has been argued (e.g., see L. G. Peck) that the horizon, N, over which extraordinary growth can be forecast is itself a function of the variance or "dependability" of the returns stream. By this reasoning, investors would project extraordinary earnings growth over only a very limited horizon for companies where the anticipated variance of the earnings stream is large. Since it can easily be shown that $\partial(P/E)/\partial N>0$ for a growth stock according to the finite-horizon model (see Malkiel, pp. 1028–29), it follows that price-earnings multiples should be negatively related to the variance term.

[2] Moreover, since the growth rate estimates collected were specifically made for only the next five years, it would seem that this model is not consistent with the data.

[3] In some models, the growth rate is assumed to decline in stages to the final "mature" growth rate of the economy. In other models, the initial and terminal growth rates are estimated on the basis of such factors as the retention rate and the rate of return on equity.

[4] The rationale for this approach and the derivation of equation (3) is contained in Malkiel. It is assumed that after N periods, the price-earnings ratios for all stocks revert to the same average condition.

[5] The closeness of the proposed linear approximation was examined by fitting a regression of the form

$$\frac{D_{j0}}{E_{j0}}\sum_{i=1}^{5}\frac{(1+g_j)^i}{(1+r)^i} + (m_s)_0\frac{(1+g_j)^5}{(1+r)^5} = A + Bg_j + C\frac{D_0}{E_0} \tag{3'}$$

Values of the parameters $(m_s)_0$ and r were chosen to be consistent with experience during the 1961–65 period. The coefficient of determination, 0.97, was so high that it seemed safe to substitute the right-hand side of (3′) for the right-hand side of (3). It should be noted, however, that this argument assumed that the horizon N is the same for all companies.

A second justification for the inclusion of a variance term in the model rests on recent theoretical work by William Sharpe, John Lintner, and Jan Mossin, extending the Markowitz portfolio selection model. In these models the market establishes "prices" for the expected return and "risk" of each security, where risk consists of the sum of the variance of that security's return and its covariances with all other returns multiplied by the number of shares. If we assume that the returns from different securities are uncorrelated with each other, however, it turns out that the price of a security should simply be a linear function of the expected return and the variance associated with the security. This suggests not only that a variance term should be included in the model but also provides some justification for the linear specification employed in this study.

The second risk measure employed in this study, an index of the conformance between the returns of each individual security and that of a market index, rests on more realistic assumptions. In Sharpe's simplification of the Markowitz model, covariances are assumed to arise because all returns depend on one or a few common factors, such as a market or industry return. For example, the returns from each security, R_i, might first be related to the returns from some index of security prices

$$(4) \quad R_i = \alpha_i + \beta_i(\text{Return to Index}) + \mu_i$$

The total risk of an asset (i.e., the scatter of the R_i around their mean), can then be decomposed into a systematic component (due to underlying relationship between R_i and the return from the market index) and a nonsystematic component, μ_i, uncorrelated with the market index. We would expect investors to prefer those securities with low or negative β_i's. Other things being equal, a stock whose movements are not highly correlated with the market will tend to reduce the variability and thus, the risk of the stock portfolio. Of course, it should be emphasized that the covariances and variances that are being valued in the market are those perceived by investors and not some "true" set.

The final risk variable employed was a leverage variable measuring the "financial risk" of a company. As Franco Modigliani and Merton Miller have shown, leverage can be expected to decrease the price-earnings multiple by increasing the riskiness of the returns of common stock relative to their expected values. With a fully adequate measure for the risk associated with the stock, leverage should play no part. Otherwise, it may serve as a useful proxy for the expected variability of the returns stream. Indeed, if other risk measures apply to the instability of the operating earnings stream before fixed charges, and thus serve as estimates of the "business risk" of the firm, a leverage term may capture the additional financial risk of the firm.

Before ending this discussion of the general model underlying the study,[6] it is worth emphasizing that the model is cast entirely in terms of expectational variables. The critical dependence of share prices on expectational variables has proved to be a major obstacle for empirical investigators. Since only historical data have been available to most researchers, it has been difficult to isolate the true effect of the various variables influencing stock prices. A simple illustration should make this clear. The model described above indicates that we should expect that a *ceteris paribus* increase in the dividend-payout ratio should increase the price-earnings multiple of the shares.[7] Suppose, however, that the past

[6] In a forthcoming publication, the authors will present a thorough and integrated model of share valuation.

[7] We must be careful, however, not to interpret a positive dividend coefficient as indicating that an individual firm can increase the price-earnings ratio of its shares by raising the dividend-payout ratio. A higher dividend (lower retention rate) may lower the future

growth rate of earnings is a very imperfect substitute for the relevant expected growth rate security purchasers anticipate.[8] The dividend payout could actually serve as an alternative proxy for expected growth.

For example, investors may take a low dividend payout ratio as a signal that the firm has many profitable investment opportunities available and that a high rate of earnings growth can be expected. In such a case, the coefficient of the payout ratio will be biased downward.[9] Without the proper expectational variables, it will be impossible to untangle the true influence of the many factors influencing the structure of price-earnings multiples. The following section will discuss the actual data employed in the study and indicate how they were collected.

II. *A Description of the Data Employed*

The principal data used in the study consist of a small number of forecasts of the long-term growth rates of earnings for 178 corporations, as of the five year-end periods from 1961 through 1965. In addition, data were collected on security analysts' estimates of "normal" earnings for the preceding year, their forecasts of next year's earnings, and their expectations about the future variability of the earnings stream. Certain historical financial data were also used to provide a contrast with the expectations data. These included past growth rates of various financial variables, past dividend-payout ratios, and a number of calculated risk proxies.[10]

The expectations data were collected from 17 investment firms, most of which were members of The Institute for Quantitative Research in Finance.[11] Of the participating firms, four were brokerage houses doing a considerable amount of investment advisory and institutional business, five were banks heavily engaged in trust management, five were mutual-fund management companies, two were pension-fund managers, and the remaining participant was an insurance company. The sample of 178 corporations was selected on the basis of data availability. Companies were included in the sample only when several investment firms made estimates of future earnings growth. Since there tended to be considerable overlap in the coverage of the security analysts for the leading industrial and utility companies, our sample tends to contain the "blue-chip" group of companies in which investment interest is centered. A detailed description of the data used in the study follows:

(a) *Normalized Earnings*

It is well known that the market does not necessarily capitalize the reported accounting earnings for a firm during the preceding year. If, for example, reported earnings are affected unfavorably by such

growth rate per share by an amount sufficient to keep the price of the shares constant. Thus, the standard dividend model of share valuation is in no way inconsistent with the result of Miller and Modigliani that dividend policy cannot effect the value of the enterprise.

[8] It may be argued that one should not put so much reliance on either past or expected growth rates to explain security prices since there is considerable evidence that earnings growth is "higgledy piggledy." I. M. D. Little and Cragg and Malkiel have shown that both historic growth rates and even the forecasts of security analysts are little related to the growth that is actually achieved. This may be true and yet security analysts may continue to estimate the worth of shares and their anticipated future returns on the basis of the anticipated growth rate of the security's earnings. As is well known from work on the term structure of interest rates, expectations need not be correct to be an important determinant of the yield curve. Surely it is an empirical question whether or not the market actually does value shares consistently with the model presented here.

[9] For a full discussion of the pitfalls involved in isolating the effect of dividend policy on share prices, see Friend and Puckett.

[10] All historical data were taken from the *COMPUSTAT* tapes made available by Standard Statistics Corporation.

[11] The Institute is a consortium of 30 investment firms, organized to promote quantitative research in finance.

nonrecurring factors as strikes or flood damage, or by a cyclical contraction, it is likely that investors apply an appropriate price-earnings multiple to the amount they consider to represent the normal earning power of the company. Indeed, one of the first jobs of a security analyst is to adjust the firm's accounting earnings to arrive at an indication of true earning power (see B. Graham, D. L. Dodd, and S. Cottle ch. 34). Thus, the price-earnings ratios that are relevant for valuation may be the ratios of prices to normalized earnings rather than ratios of prices to reported earnings for the preceding accounting period. These normalized earnings are estimated to be the earnings that would obtain at a normal level of economic activity if the company were experiencing normal operations—that is, operations not affected by such nonrecurring items as strikes, natural disasters, and so forth. The normalized-earnings figures used in the present study were averages of estimates supplied by two of the participating firms.

(b) *Future Long-term and Short-term Growth Rates*

As was mentioned above, several theoretical models of stock valuation have all focused on the expected growth rates of earnings and dividends as a central explanatory variable. Most previous empirical studies, however, were forced to rely on past growth rates as a proxy for future growth rates. One of the major purposes of the present study was to ascertain whether the estimates of future growth rates from several securities firms can enable us to obtain a more satisfactory explanation for the structure of share prices.

In order to contrast the use of historical and expected growth rates, we first tried to find those historical growth rates that showed the closest correlation with market price-earnings multiples. Forty alternative growth rates were tried. These growth rates differed with respect to the period covered, the method of calculation, and the financial data upon which the growth rate was estimated. From the forty candidate growth rates, the following three were either clearly superior or, at least, no worse than any of the others. These were 1) the ten-year growth rate of earnings per share calculated as the geometric mean of first ratios, 2) the ten-year growth rate of cash earnings per share (i.e., earnings plus noncash charges) calculated as the geometric mean of first ratios, and 3) the ten-year growth rate of cash earnings plus taxes calculated from a regression of the logarithms of the earnings on time. The growth rate of cash earnings was slightly better than the other two in most of the five years studied, and was used in the regressions reported in this paper.

The expected growth rates were estimated by nine securities firms.[12] Each growth rate figure was reported as an average annual rate of growth of earnings per share expected to occur over the next five years. The figures used in the study were averages of the nine predictors.

In addition to these expectations of long-term growth rates, we also collected estimates of the following year's earnings from eleven securities firms.[13] We found, somewhat to our surprise, that the implicit forecasts of short-term (one-year) growth were not highly correlated with the long-term anticipations and we were able to use both sets of data in some of the empirical work presented later.

Obviously these expected growth rates are not the expectations of a wide cross-section of the buyers and sellers in the market. These expectations were formed,

[12] It should be noted that not all firms provided growth-rate estimates for each of the companies used in the sample during each of the five years, 1961–65.

[13] Three of these eleven firms also supplied long-term forecasts.

however, by professional security analysts for securities firms or for large institutional investors who are important participants in the market. Moreover, in many cases, these expectations were made to be provided to other investors whose own expectations may be influenced by their advisors. Finally, we should note that these expectations are not limited to published information. The security analysts involved frequently visit the companies they follow and discuss the company's prospects with its executives. Insofar as other security analysts follow the same sort of procedures as our participating firms, the growth-rate estimates of other institutional investors and securities firms may resemble those we have collected. Consequently, these predictions may well serve as acceptable proxies for market expectations and they surely seem worthy of detailed analysis.

(c) *Dividend Payout*

The measurement of the dividend-payout ratio also presents problems. If we simply take the ratio of dividends to earnings, short-run disturbances to reported earnings that do not produce equiproportional changes in dividends can make calculated payouts differ considerably from target or normal payout ratios. For this reason we chose two alternative methods of calculating the dividend payout. The first method was simply to divide the dividend by normalized rather than reported earnings. The second method, used in the regressions where only historic data were employed, was to average the actual payout ratios over the preceding seven years.

(d) *Risk Variables*

Several types of expectational risk variables were introduced to serve as proxies for the anticipated variance of individual security returns. We included such variables as the standard deviation of the forecasts of security firms, various types of subjective quality ratings, and an index of the expected instability of future earnings. These risk proxies all turned out to be highly correlated with each other and only the one most useful in explaining earnings multiples, the instability index, has been included in the regressions reported in this paper. This variable was collected from one of our participating firms and represented a measure of the past variability of earnings (around trend) adjusted by the security analyst to indicate anticipated future variability.

In order to contrast the use of expectations data with historical data, a number of risk proxies were calculated on the basis of the financial records of each company. These included statistics measuring the variance of past earnings and of other financial data, a leverage variable, and the conformance between returns of each individual security and that of a market index. The index of market conformance was obtained by estimating the slope, β_i, of a regression of the annual returns of each security on the annual returns from the Standard and Poor's Composite Index. Ten years of data were employed in obtaining the estimate. The most useful historic risk proxies for our present purposes were the semideviation of earnings around trend, the index of market conformance, and the leverage variable. In Table 1 we summarize the variables employed in the regressions.

Before turning to the regression results, a problem concerning the timing of the availability of the expectations and hisrorical data should be mentioned. Our study tries to explain differences among price-earnings multiples for a cross-section of securities as of December 31 in each of five years. While normal earnings per share (and expected growth rates) were estimated and, therefore, available at the

end of each year, actual earnings per share for the 12 months to December 31 are not generally known until some time after the close of the year. Thus, the actual P/E ratios and the historic growth rates calculated to the end of the year, which we employed in the regressions estimated from historic data, were not available to investors on the dates for which equations were estimated, although rather close estimates of the earnings necessary for the calculations are usually well known by that time. In order to test whether our results might be strongly influenced by, in effect, assuming perfect foresight by the market regarding current-year's earnings, we performed an alternative set of runs using the most recent publicly available 12-months' earnings to calculate P/E ratios and historic growth rates. Since the regression results from the alternative set of runs were almost identical to those reported here, it seems safe to conclude that our assumptions regarding the timing of the availability of historic data had little influence on the results.

TABLE 1—VARIABLES USED IN VALUATION STUDY

P	End-of-year market price per share
D	Total dividends paid per share (adjusted to number of shares outstanding at year end)
E	Reported earnings per share (adjusted to exclude nonrecurring items)
$\overline{D/E}$	Average dividend-payout ratio over past 7 years
$\overline{NE}$	Average "normalized" earnings estimates of security analysts
$\bar{g}_p$	Average predicted future long-term growth rate of earnings per share, measured as an annual percentage rate of growth
g_H	Historic (10-year) growth rate of (cash) earnings per share measured as an annual percentage rate of growth
I_p	Predicted instability index of the future earnings stream
β	The slope of a regression of the annual returns from a company's shares on the annual returns from the market index
$I_{H,1}$	Calculated instability index of the historic earnings stream (semideviation of earnings around trend)
$I_{H,2}$	Calculated instability index of the historic operating earnings streams (semideviation of earnings plus financial fixed charges around trend)
$\bar{E}_{t+1}$	Average predicted earnings per share for the next year
$\frac{F}{E+F}$	Leverage variable (the ratio of fixed charges to earnings plus fixed charges)

III. *Regression Results*

In this section we first present a comparison of the regression results for equations including comparable historic and expectational variables. Then, the results for the most satisfactory expectational equations are shown and the stability of the coefficients over time is examined.

(a) *Comparison of Regressions Using Historical and Expectational Variables*

In Table 2 the results of regressions using only three variables calculated from readily available historical data are compared with regressions employing comparable expectations data.[14] In panel A of Table 2, the price-earnings multiple is regressed on the historic ten-year growth rate of cash earnings (calculated as the geometric mean of first ratios), the dividend-payout ratio (averaged over the preceding seven years), and an instability index of earnings (calculated as the semideviation from a regression of earnings over the past ten years). It will be noted that generally about half of the variance in price-earnings multiples is explained by the regressions. The growth-rate variable is highly significant in each of the years covered. The calculated payout and risk

[14] It will be noted that the sample size for each regression was usually less than the total sample of 178 companies. Companies had to be dropped from the sample whenever historic or expectational data were unavailable or could not be computed. In addition whenever a company's calculated historic growth rate was negative, the firm was dropped from the sample. This was done to make the regressions based on historic data as comparable as possible to those based on expectations data, where no negative growth rates were projected.

TABLE 2—COMPARISON OF REGRESSIONS USING HISTORICAL AND EXPECTATIONAL VARIABLES

A. REGRESSION RESULTS: HISTORIC VARIABLES

$$P/E = a_0 + a_1 g_H + a_2 \overline{D}/\overline{E} + a_3 I_{H,1}$$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2	F
1961	13.65	+1.87	−.26	−.65	.50	51.27
		(.17)	(6.14)	(1.37)		(3; 156)
		10.72	−.04	−.47		
1962	8.92	+1.06	+6.90	−.77	.45	44.78
		(.10)	(3.28)	(.68)		(3; 163)
		10.90	2.10	−1.14		
1963	9.39	+1.33	+5.22	−.96	.49	51.31
		(.12)	(3.73)	(.81)		(3; 161)
		11.29	1.40	−1.19		
1964	10.88	+.95	+4.85	−.69	.36	32.16
		(.11)	(3.52)	(.71)		(3; 170)
		8.65	1.38	−.96		
1965	5.74	+1.52	+6.64	+.35	.65	98.65
		(.10)	(3.55)	(.77)		(3; 162)
		15.23	1.87	.46		

B. REGRESSION RESULTS: COMPARABLE EXPECTATIONS VARIABLES

$$P/E = a_0 + a_1 \bar{g}_p + a_2 D/\overline{NE} + a_3 I_p$$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2	F
1961	4.73	+3.28	+2.05	−.82	.70	89.34
		(.23)	(4.33)	(.75)		(3; 115)
		14.47	.47	−1.09		
1962	11.06	+1.75	+.78	−1.61	.70	133.33
		(.13)	(2.47)	(.39)		(3; 174)
		13.99	.31	−4.11		
1963	2.94	+2.55	7.62	−.27	.75	174.51
		(.13)	(2.58)	(.39)		(3; 174)
		19.67	2.95	−.69		
1964	6.71	+2.05	+5.33	−.89	.75	168.46
		(.11)	(2.17)	(.36)		(3; 170)
		18.24	2.44	−2.48		
1965	.96	+2.74	+5.01	−.35	.85	317.52
		(.10)	(2.05)	(.30)		(3; 171)
		26.50	2.44	−1.14		

Note: Numbers in parentheses below coefficients are standard errors and numbers below parentheses are t-values. Numbers below the F-values are degrees of freedom.

measures usually have their expected signs but are not significant.[15]

In panel B of Table 2, the average growth rates and other expectational variables collected from the participating firms are used to explain price-earnings multiples. All coefficients have their expected signs. Moreover, the fits are very close for cross-sectional empirical work and are much better than those obtained with the historical data. About three quarters of the variability of price-earnings ratios is explained by the regressions. We should also mention that better fits were obtained by using the average growth rates of all predictors than by employing forecasts of a single analyst. This suggests that our survey was useful in getting closer to what might be considered the expectations of a "representative" investor.

(b) *Regression Results Employing a Covariance Risk Measure*

In Table 3 we present regression results employing a covariance risk measure. It will be noted that β, the index of market conformance, has the right sign in all cases except for the 1961 regression employing expectations data. Although it is significant in only two of the five years, the general consistency of the signs would suggest that market values do tend to reflect measures of past covariance with the market. It is also interesting that β had a particularly strong influence on price-earnings ratios at the end of 1962, following a large decline in stock prices. It would appear that investors particularly favor securities that tend to move relatively independently of the market during periods when the memory of sharply falling stock prices is clearly in mind.

Comparing Tables 2 and 3, the t-values associated with β tend to be slightly higher than those associated with either of the two previous risk variables.[16] When a variable measuring expected short-term growth is introduced, however, the predicted instability index tends to be somewhat superior, being "significant" in four out of the five years (see Table 5). The variables β and I_p cannot be used together in the same regression, because the two variables are highly correlated, and both become insignificant.[17]

(c) *Regression Results Employing a Combination of Expectations and Historic Data*

In Table 4, we present regression results involving a combination of expectations and historic data. The price-normalized earnings ratio is employed as the dependent variable. Independent expectational variables include anticipations of short- and long-term growth, and the dividend payout expressed as a percent of normalized earnings. Historic variables were an instability index and a leverage variable. In these regressions, the instability index was calculated from a time-series of earnings plus fixed charges. This measure should represent the instability of operating earnings and may serve as an acceptable proxy for business risk. We also included a leverage variable, which should indicate the additional financial risk borne

[15] As noted above, the positive sign on the dividend coefficient should not be interpreted as evidence that dividend policy can affect the value of the shares. This coefficient indicates only that a *ceteris paribus* change in dividend payout will increase the price of the shares. What the famous "dividend-irrelevancy" theorem of Modigliani and Miller says is that an increase in dividend payout (holding the firm's investment constant) will tend to reduce the growth rate of earnings per share since new shares will now have to be sold to make up for the extra funds paid out in dividends. A positive dividend coefficient is thus in no way inconsistent with the dividend-irrelevancy theorem.

[16] While it should be noted that these comparisons are based on regressions using somewhat different numbers of observations, the conclusions presented hold also for comparisons based on the smaller sample of companies for which all data were available.

[17] Correlation coefficients between β and I_p during the period studied are approximately 0.60.

TABLE 3—REGRESSION RESULTS EMPLOYING A COVARIANCE RISK MEASURE

A. HISTORIC VARIABLES AND COVARIANCE MEASURE

$$P/E = a_0 + a_1 g_H + a_2 \overline{D}/\overline{E} + a_3 \beta$$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2	F
1961	15.52	1.82 (0.17) 10.54	−1.75 (6.14) −0.29	−1.53 (1.34) −1.15	.49	52.60 (3; 161)
1962	12.42	1.02 (0.09) 11.38	4.28 (2.94) 1.46	−2.87 (0.60) −4.76	.54	65.86 (3; 169)
1963	9.20	1.28 (0.11) 11.19	6.84 (3.67) 1.87	−1.21 (0.88) −1.38	.48	51.69 (3; 168)
1964	14.37	0.96 (0.10) 9.36	3.29 (3.18) 1.03	−3.54 (0.72) −4.92	.44	44.76 (3; 173)
1965	7.47	1.52 (0.10) 15.30	5.58 (3.34) 1.67	−0.95 (0.79) −1.20	.64	99.49 (3; 165)

B. COMPARABLE EXPECTATIONS VARIABLES AND COVARIANCE MEASURE

$$P/E = a_0 + a_1 \hat{g}_p + a_2 D/\overline{NE} + a_3 \beta$$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2	F
1961	3.63	3.29 (0.19) 17.20	3.24 (4.47) 0.73	0.97 (1.09) 0.89	.74	132.82 (3; 140)
1962	9.79	1.87 (0.11) 16.88	2.25 (2.23) 1.01	−2.65 (0.47) −5.69	.72	148.29 (3; 173)
1963	3.47	2.57 (0.12) 21.38	7.17 (2.47) 2.90	−0.84 (0.61) −1.37	.75	176.82 (3; 174)
1964	6.16	2.10 (0.10) 21.40	5.87 (2.04) 2.88	−1.41 (0.53) −2.67	.76	184.63 (3; 173)
1965	0.25	2.86 (0.10) 29.14	5.01 (2.00) 2.50	−0.47 (0.49) −0.96	.86	352.19 (3; 172)

Note: Numbers in parentheses below coefficients are standard errors and numbers below parentheses are t-values. Numbers below the F-values are degrees of freedom.

by the shareholders. The specific measure employed was the ratio of fixed charges per share to earnings plus fixed charges per share.[18] In addition, a dummy variable was included that took the value unity for utility companies and zero for industrials. This variable was introduced to account

[18] For a discussion of the problems involved in using the debt-equity ratio itself, see A. Barges and R. Wippern.

TABLE 4—REGRESSION RESULTS EMPLOYING A COMBINATION OF EXPECTATIONS AND HISTORIC DATA

$$P/\overline{NE}=a_0+a_1\bar{g}_p+a_2\bar{E}_{t+1}/\overline{NE}+a_3D/\overline{NE}+a_4F/(E+F)+a_5Dum+a_6I_{H,2}$$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	R^2	F
1961	−41.19	+2.88	+44.88	+5.53	−12.34	+1.79	−4.93	.85	102.98
		(.20)	(5.24)	(4.53)	(4.06)	(1.69)	(9.21)		(6;106)
		14.07	8.57	1.22	−3.04	1.05	− .54		
1962	−1.41	+1.68	+9.89	+2.60	−7.53	+4.46	−7.69	.78	74.04
		(.13)	(2.72)	(2.50)	(2.07)	(.92)	(4.75)		(6;129)
		13.16	3.63	1.04	−3.65	4.87	−1.62		
1963	−12.94	+2.41	+15.29	+8.96	−6.20	+.71	−5.70	.81	90.72
		(.14)	(2.99)	(2.79)	(2.33)	(1.04)	(5.33)		(6;129)
		17.12	5.11	3.21	−2.66	.69	−1.07		
1964	−10.91	+1.89	+14.31	+7.70	−3.39	+3.62	+4.59	.80	83.42
		(.12)	(2.02)	(2.45)	(2.21)	(.94)	(5.28)		(6;128)
		15.65	7.09	3.14	−1.53	3.86	(.87)		
1965	−15.55	+2.64	+20.05	−2.04	−7.81	+2.64	−17.59	.84	118.41
		(.14)	(1.99)	(3.01)	(2.61)	(1.12)	(6.33)		(6;128)
		18.69	10.09	−.68	−2.99	2.37	−2.78		

Note: Numbers in parentheses below coefficients are standard errors and numbers below parentheses are t-values. Numbers below the F-values are degrees of freedom.

for differences in risk between the two classes of companies not captured by our other risk variables.

As can be seen from the table, the combination of historical and expectational variables works remarkably well in accounting for the structure of share prices. Most significant were the coefficients of the short- and long-term growth rates. It should be noted that while the coefficient of the "operating-risk" variable (the semi-deviation of earnings plus fixed charges around trend) usually was not statistically significant and had the "wrong" sign in 1964, the coefficient of the financial-risk variable (our measure of leverage) always had the "correct" sign and was significant in all but one year. This provides support for the Modigliani-Miller proposition that the required rate of return on equity should be an increasing function of leverage.

(d) *Regression Results Employing Expectations Data Alone*

In Table 5 we present additional regression results for the equations employing only expectations variables. The price-normalized earnings ratio is the dependent variable. Independent variables include expectations of short- and long-term growth, the dividend-payout ratio, and the expected instability index.[19]

We find that the long-term growth variable contributes most to an explanation of the structure of earnings multiples. The growth coefficient has a t-value over 13 in every year. The coefficient of short-term growth ($\bar{E}_{t+1}/\overline{NE}$) is also positive and highly significant. The coefficients of the payout ratio and the risk proxy are positive and negative, respectively, as ex-

[19] Fortunately, the correlations between the independent variables tended to be relatively low in all years. A sample correlation matrix (for the 1964 data) is presented below

	$\bar{g}_p$	$\bar{E}_{t+1}/\overline{NE}$	I_p	$D/\overline{NE}$
$\bar{g}_p$	1.00			
$\bar{E}_{t+1}/\overline{NE}$	.28	1.00		
I_p	− .32	.09	1.00	
$D/\overline{NE}$	− .34	− .07	− .37	1.00

TABLE 5—REGRESSION RESULTS: EMPLOYING EXPECTATIONS DATA

	$P/\overline{NE}=a_0+a_1\bar{g}_p+a_2\bar{E}_{t+1}/\overline{NE}+a_3D/\overline{NE}+a_4I_p$							$R_{t+1}=a+b\left[\frac{P/\overline{NE}-\hat{P}/\overline{NE}}{\hat{P}/\overline{NE}}\right]$		
Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	R^2	F	$\hat{b}$	R^2	F
1961	−27.96	+2.91 (.21) 13.56	+31.78 (5.76) 5.51	+4.57 (3.96) 1.15	−.58 (.70) −.83	.77	80.39 (4,96)	−.25 (.08) −3.08	.09	9.47 (1;99)
1962	+3.42	+1.61 (.12) 13.05	+6.88 (2.87) 2.40	+3.21 (2.32) 1.39	−2.20 (.41) −5.44	.79	129.14 (4,138)	.21 (.11) 1.93	.03	3.73 (1;141)
1963	−11.33	+2.29 (.14) 16.30	+15.11 (2.82) 5.35	+8.11 (2.70) 3.01	−1.14 (.39) −2.88	.80	139.82 (4,137)	−.20 (.08) −2.55	.04	6.48 (1;140)
1964	−9.29	+1.87 (.14) 13.05	+15.20 (1.94) 7.83	+7.03 (2.40) 2.92	−1.13 (.41) −2.75	.78	120.00 (4,134)	−.00 (.15) −.00	.00	.00 (1;137)
1965	−11.15	+2.42 (.12) 19.59	+13.78 (1.85) 7.46	+4.22) (2.34) 1.81	−.81 (.38) −2.14	.83	162.21 (4,136)	—.01 (.10) −.11	.00	.01 (1;139)

Note: Numbers in parentheses below coefficients are standard errors and numbers below parentheses are t-values. Numbers below the F-values are degrees of freedom.

pected, and are usually significant. While Tables 4 and 5 are not comparable because of different degrees of freedom, the regressions in Table 5 tend to produce slightly better fits adjusted for degrees of freedom.

It might be argued that the expectations data used as independent variables in the valuation equation may strongly reflect the $P/\overline{NE}$ ratio and, thus, we are in effect including the same variable on both sides of the valuation equation. The growth rates that we have collected are "supposedly" independent of market prices. The security analysts who have furnished the data claim that these estimates are ones that they use to calculate an "intrinsic" value of the shares, which is then compared with actual market prices in arriving at purchase or sale recommendations. In point of fact, however, the forecasted growth rates may still be strongly influenced by the market earnings multiples themselves.

Even if the anticipations data are strongly influenced by current market prices, however, this should not interfere with the basic purpose of this paper, which is to gain an understanding of the structure of share prices. The point is that the anticipations we have collected may simply be the security analysts' estimates of what the "average opinion" will continue to believe the reasonable expectations will be. The point is, of course, the familiar one about the Keynes beauty contest where the rational contestant would not pick those girls that he himself found prettiest, nor even those he deemed most likely to catch the fancy of the other contestants, but rather those that he anticipated the other contestants would believe the average opinion would consider prettiest.

Thus, if the $P/\overline{NE}$ ratio rises, and the security analyst believes that such a rise will continue to be justified by the average opinion, he may simply adjust his antici-

pated growth rate to a level that would justify the earnings multiple. In any case, what our valuation equation will measure is the relationship between growth rates and price-earnings multiples that security analysts believe the average opinion will continue to justify. Even in this event, our empirical results should still be useful in explaining and describing the structure of share prices at any given time.

(e) *Changes in the Valuation Relationship Over Time*

It is of some interest to examine whether the coefficients of the valuation equations are the same in each year or whether they change. This is of considerable importance to those who wish to use valuation equations in connection with assigned values of the independent variables to estimate the intrinsic worth of a security. Constancy of the relationship is also important if a firm is to seek to follow policies that will maximize the value of its shares. On the other hand, there is nothing in the theory of valuation to indicate that the equation need be constant over time.

An inspection of Table 5 indicates that the coefficients of our equation change considerably from year to year and in a manner that is consistent with the changing standards of value in vogue at the time. At the end of 1961 "growth stocks" were in high favor, and it is not surprising to find that the coefficient of the growth rate (2.91) is highest in this year. During 1962, however, there was a conspicuous change in the structure of share prices that was popularly called "the revaluation of growth stocks." This revaluation is reflected in the decline of the growth-rate coefficient for 1962 to 1.61, its lowest value for any of the five years. A similar set of observations can be made for the coefficient of the short-term growth rate ($\bar{E}_{t+1}/\overline{NE}$). On the other hand, the risk index has its most negative influence on earnings multiples in 1962, whereas the coefficient was smallest in 1961, and, while negative, it was not significantly different from zero.

In actually testing whether the coefficients of the valuation equation were the same over time, it had to be recognized that the residuals in different years might not be independent. Indeed, it is shown in the bottom panel of Table 6, which we will discuss below, that the residuals are fairly highly correlated. As a result, Arnold Zellner's seemingly unrelated regression version of Aitken's generalized least-squares model is appropriate, although it had to be modified to take account of the fact that we did not have observations for all corporations in all years.[20] Using this procedure, the hypothesis that the coefficients are the same in each year was rejected beyond the .0001 level.

IV. *Use of the Valuation Model for Security Selection*

One of the most intriguing questions concerning empirical valuation models is whether they can be used to aid investors in security selection. The empirical valuation equation shows us, at a moment in time, the average way in which variables such as growth, payout, and risk influence market price-earnings multiples. Given the values of these variables applicable to any specific company, we can compute an estimated normal price-earnings ratio based on the empirical valuation equation. It has been suggested that securities may be selected by comparing the actual market price-earnings ratio with the normal

[20] In using this procedure, the covariance matrix of the disturbances was estimated from the single-equation regression residuals. This procedure also produced more efficient estimates of the coefficients of the individual equations. Since these differed but little from those shown in Table 5, and had the same implications, we shall not present them here. The test reported is an F-test (asymptotically), which uses the vectors of independent and dependent variables, following transformation, in the usual way.

TABLE 6—ANALYSIS OF LACK OF FORECASTING SUCCESS

Description	Coefficient of Determination Residuals against 1964 Return	F-Value (and Degrees of Freedom)
1963 Valuation equation with 1963 predictions	.04	6.48 (1; 140)
1964 Valuation equation with 1963 data. (Assume that next year's valuation relationship is known.)	.08	12.15 (1; 140)
1963 Valuation equation with realized growth rates. (Assumes perfect foresight regarding future long-term growth and next year's earnings.)	.12	18.140) (1; 14
1963 Valuation equation with 1964 predictions. (Assumes perfect foresight regarding market expectations next year.)	.24	41.75 (1; 140)

Correlations of Residuals over Years

Description	Coefficient of Determination
1962 vs. 1961	.46
1963 vs. 1962	.24
1964 vs. 1963	.13
1965 vs. 1964	.35

multiple predicted by the valuation equation. If the actual earnings multiple is greater (less) than the normal earnings multiple, we designate the security as "overpriced" ("underpriced") and recommend sale (purchase). Such a procedure was employed by Whitbeck and Kisor, who claimed that an underpriced group of securities selected by the above procedure consistently outperformed an overpriced group during the early 1960's.

Of course, even on a priori grounds, it is possible to think of many reasons why such a procedure would prove fruitless. For example, if high P/E (high growth rate) stocks tended to be overpriced during one particular period, the estimated growth-rate coefficient will be larger (by assumption) than that which is warranted. However, the recommended procedure will not indicate that high P/E stocks are overpriced because normal market-determined earnings multiples for these securities will themselves be higher than is warranted. Nevertheless, in view of the positive results reported by Whitbeck and Kisor, it would seem desirable to attempt to replicate their experiment with our data.

The results of some of our experiments are shown in the right-hand columns of Table 5. We measured the degree of over- or underpricing as the ratio of the residual from the prediction equation to the predicted earnings multiple, i.e., $[(P/\overline{NE}-\hat{P}/\overline{NE})/(\hat{P}/\overline{NE})]$. A percentage measure was chosen in view of the considerable variance in actual earnings multiples. If the model is useful in measuring underpricing, then underpriced securities, according to this criterion, ought to "outperform" overpriced issues over some subsequent period. We picked one year as the appropriate horizon and measured

subsequent returns, in the normal manner, as

$$R_{t+1} = \frac{P_{t+1} - P_t + D_{t+1}}{P_t} \tag{5}$$

If the empirical valuation model is successful in selecting securities for purchase, the percentage residual (degree of overvaluation) from the valuation equation ought to be negatively related to these subsequent returns. As the table indicates, in only three of the five years for which this experiment was performed was the relationship negative, and the degree of association was extremely low. In the other two years, there was either a positive or zero relationship. Supplementary tests conducted by industry and other groupings produced similar results. It should also be noted that the residuals from the equations employing historical data and from equations combining historical and expectational data were no more successful in predicting subsequent performance. Moreover, these results were unaltered when the subsequent returns were measured over alternative time periods such as one quarter ahead or two or more years ahead.

In Table 6 some statistics are presented which may be helpful in interpreting the reason for our predictive failures. We note that using the 1963 valuation equation as an example, the percentage degree of under- or overpricing is not highly correlated with subsequent returns. The coefficient of determination is only .04. It is possible, however, to isolate four reasons for our lack of forecasting success.

1) The first reason is that the valuation relationship changes over time. We might be unable to select truly underpriced securities because by the next year (the end of the horizon period) the norms of valuation have been significantly altered. Thus, what was cheap on the basis of 1963's relationship may no longer represent good value on the basis of the 1964 relationship. To test how important this factor might be, we performed the following experiment: We assumed that investors knew at the end of 1963 exactly what the market valuation relationship would be in 1964, i.e., we assumed perfect foresight regarding next year's valuation equation. Then, on the basis of the 1964 valuation equation, we utilized the 1963 data to calculate warranted $P/\overline{NE}$ multiples, which could then be compared with actual multiples to determine whether each security was appropriately priced. Correlating the percentage residuals with subsequent returns, we found that the coefficient of determination doubled, 8 percent of the variance in subsequent returns was explained.

2) A second reason for lack of success might be the quality of the expectations data employed. As was indicated in our 1968 article several of the growth-rate forecasts used in the present study were in fact shown to be rather poor predictors of realized earnings growth. To determine how much better off we would be with more accurate forecasts, we assumed perfect foresight regarding the future long-term growth rate of the company and regarding the next year's anticipated earnings. Thus, the 1963 empirical valuation equation was used to determine normal value, but in place of the variable $\bar{E}_{64}/\overline{NE}_{63}$ we substituted the variable $E_{\text{actual }64}/\overline{NE}_{63}$, and in place of $\bar{g}_p$ we substituted the realized long-term growth rate through the end of 1966. Using these realized data to determine warranted price-earnings multiples, the percentage residuals therefrom were correlated with future returns. As expected, an even greater improvement in forecasting future returns was found. The R^2 rises to .12.

3) As a further experiment, perfect foresight was assumed not regarding the

actual rate of growth of earnings but rather regarding what the market expectations of growth would be next year. Calculating the degree of overpricing as before, we find a much greater improvement in prediction of future returns, 24 percent of the variability of future returns is explained, compared with 4 percent in the original experiment. We conclude that if one wants to explain returns over a one-year horizon it is far more important to know what the market will think the growth rate of earnings will be next year rather than to know the realized long-term growth rate. Of course this observation brings us back to Keynes' newspaper contest again. What matters is not one's personal criteria of beauty but what the average opinion will expect the average opinion to think is beautiful at the close of the contest.

4) A final source of error is that the valuation model does not capture all the significant determinants of value for each individual company. Despite our success in accounting for approximately 80 percent of the variance in market price-earnings multiples, there are likely to be special features applicable to many individual companies that cannot be captured quantitatively. For example, it turned out that the stock of Reynolds Tobacco always appeared to be underpriced. The reason for this is, of course, not difficult to conjecture. There is a risk of government sanctions against the tobacco industry, which weighs heavily in the minds of investors, but which is not related to the instability measure of Reynolds' earnings we have employed.

To indicate how important this problem of omitted variables might be, the residuals from our valuation equations from year to year were correlated. If certain factors specific to individual companies are consistently missing, the residuals from the valuation equations can be expected to be positively correlated over time. As the bottom half of Table 6 indicates, the residuals are significantly correlated over time. Thus, despite our success with expectations data in estimating a valuation equation which has far more explanatory ability than those based on historic information, it is clear that certain systematic valuation factors are still missing from the analysis.[21] Consequently, it cannot be said that all deviations of actual from predicted price-earnings ratios are simply manifestations of temporary over- or underpricing.

V. *Concluding Comments*

We have demonstrated that it is possible to explain, for several successive years, a large percentage of the variability in market price-earnings ratios with the variables included in this study and the specification suggested by the very simple model in Section I. The analysis was not successful, however, in isolating underpriced securities that might be expected to have above-average future returns. Needless to say, there are many additional factors that should be considered in a full valuation study. While it does not seem likely that this further work will provide direct answers to the problem of security selection, it may well shed further light on the logic of market valuations.

REFERENCES

T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, New York 1958.

F. D. Arditti, "Risk and the Required Return

[21] It would be possible, of course, to incorporate some sort of firm-effect variable to capture the systematic portion of the under- or overpricing. Such a procedure was attempted. As a firm effect we utilized the difference between the last year's actual $P/\overline{NE}$ multiple and and $\hat{P}/\overline{NE}$, the predicted earnings multiple. As might be expected, this procedure served to improve the goodness of fit substantially, but it did not affect the magnitude of the other regression coefficients. Unfortunately, however, it was not successful in improving the forecasts of future returns.

on Equity," *J. Finance*, Mar. 1967, *22*, 19–36.

A. Barges, *The Effect of Capital Structure on the Cost of Capital*, Englewood Cliffs 1963.

H. Benishay, "Variability in Earnings—Price Ratios of Corporate Equities," *Amer. Econ. Rev.*, Mar. 1961, *51*, 81–94.

R. S. Bower and D. H. Bower, "Risk and the Valuation of Common Stock," *J. Polit. Econ.*, May-June 1969, *77*, 349–62.

K. J. Cohen and J. A. Pogue, "An Empirical Evaluation of Alternative Portfolio-Selection Models," *J. Business*, Apr. 1967, *40*, 166–93.

J. G. Cragg and B. G. Malkiel, "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings," *J. Finance*, Mar. 1968, *23*, 67–84.

G. R. Fisher, "Some Factors Influencing Share Prices," *Econ. J.*, Mar. 1961, *71*, 121–41.

I. Friend and M. Puckett, "Dividends and Stock Prices," *Amer. Econ. Rev.*, Sept. 1964, *54*, 656–82.

M. J. Gordon, *The Investment, Financing, and Valuation of the Corporation*, Homewood 1962.

B. Graham, D. L. Dodd, and S. Cottle, *Security Analysis, Principles and Technique*, 4th ed., New York 1962.

S. E. Guild, *Stock Growth and Discount Tables*, Boston 1931.

C. C. Holt, "The Influence of Growth Duration on Share Prices," *J. Finance*, Sept. 1962, *17*, 464–75.

F. C. Jen, "Common Stock Valuation: An Empirical Investigation of the Finite Horizon Hypothesis," paper presented to the meeting of the Amer. Statist. Ass., Sept. 1965.

M. Kisor, Jr. and A. Feuerstein, "Towards a Valuation Model Employing Historical Constructs as Proxies for Analysts' Expectations," mimeo. paper, May 1967, New York.

M. Kisor, Jr. and S. Levine, "Simulated Results from a Valuation Model Using Objectively Determined Proxies for Analysts' Expectations," mimeo. paper, New York, July 1968.

J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.*, Feb. 1965, *47*, 13–37.

I. M. D. Little, "Higgledy Piggledy Growth," *Oxford Inst. Statist. Bull.*, Nov. 1962, *24*, 387–412.

B. G. Malkiel, "Equity Yields, Growth, and the Structure of Share Prices," *Amer. Econ. Rev.*, Dec. 1963, *53*, 1004–31.

H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, New York 1959.

M. H. Miller and F. Modigliani, "Dividend Policy, Growth and the Valuation of Shares," *J. Business*, Oct. 1961, *34*, 411–33.

F. Modigliani and M. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

J. Mossin, "Equilibrium in a Capital Asset Market," *Econometrica*, Oct. 1966, *34*, 768–83.

R. Ortner, "An Estimate of the Time Horizon and Expected Yield for a Selected Group of Common Shares, 1935–1955," *Int. Econ. Rev.*, May 1961, *2*, 179–98.

L. G. Peck, "Evaluation of Electric Utility Stock Prices," unpublished paper, Goldman, Sachs & Co., New York 1969.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan.-Apr. 1964, *32*, 122–36.

W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *J. Finance*, Sept. 1964, *19*, 425–42.

P. F. Wendt, "Current Growth Stock Valuation Methods," *Financial Analysts J.*, Mar.-Apr. 1965, *21*, 91–105.

V. Whitbeck and M. Kisor, "A New Tool in Investment Decision Making," *Financial Analysts J.*, May-June 1963, *19*, 55–62.

W. P. Whitlock, "A Practical Present Value Model," paper presented to the June 1967 seminar of the Institute for Quantitative Research in Finance.

J. B. Williams, *The Theory of Investment Value*, Cambridge 1938.

R. Wippern, "Financial Structure and the Value of the Firm," *J. Finance*, Dec. 1966, *21*, 615–33.

A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Ass.*, June 1962, *57*, 348–68.

Export Instability and Economic Structure

By Benton F. Massell*

This paper examines the relationship between instability in the value of exports and a set of variables that help characterize a country's economic structure. For each of 55 countries an index of the instability of merchandise export receipts is constructed employing data for 1950–66. Regression analysis is used to explain intercountry differences in export instability in terms of nine structural variables.

During the past two decades, economists have expressed concern with export instability, principally as it affects less-developed countries (*LDC*s). It is commonly thought that fluctuations in *LDC* export earnings generate domestic instability (with a consequent welfare loss), complicate the task of development planning, and reduce the efficiency with which investment resources are used.[1] Although long-term growth has been a major objective of *LDC* policy, short-run fluctuations also appear to receive considerable weight in the determination of policy choices. One suspects that efforts to diversify exports and to shift resources from primary production to manufacturing are designed, at least in part, to decrease export instability.[2]

To the extent that export instability is influenced by factors within the control of economic planners—or is subject to influence by international agencies—it is relevant to estimate the relationship between these factors and instability. Six of the explanatory variables used in the analysis are subject to manipulation by policymakers: commodity concentration, geographic concentration, specialization on food, specialization on raw materials, export market share, and domestic consumption of exported goods.[3] Three additional explanatory variables—size of the export sector, per capita income, and a dummy variable to distinguish between developed and less-developed countries—cannot be considered as policy instruments. However these three variables are included in the regression to avoid specification bias and to improve the explanatory power of the model.[4] The final formulation of the regression equation contains five explanatory variables that explain 46

* Professor, Food Research Institute, Stanford University. This paper presents the results of research initially undertaken for the 1968 Summer Research Project of the Agency for International Development, and continued under a grant from the Center for Research in International Studies at Stanford University. In formulating and conducting this research I profited from productive discussions with, and useful suggestions from, William O. Jones, Richard N. Cooper, Scott Pearson, Peter Timmer, and Michael Roemer. Donald Keesing provided especially provocative and constructive comments on an earlier draft. The referee's valuable comments aided in making final revisions of the study. Finally, thanks are due to Adele Palmer, William Brandt, Catherine Whittemore, and Henry Moore for very competent research assistance.

[1] For a discussion of the impact of export instability on *LDC* economies, see the forthcoming paper by Massell, S. R. Pearson, and J. B. Fitch.

[2] This is evidenced by *LDC* concern with supplementary and compensatory finance.

[3] These variables are not subject to policy manipulation in the short run but can be viewed as long-run policy variables. Just as policymakers are concerned with shifting resources into higher productivity uses, so might they be concerned with allocating resources to achieve either greater diversification among goods or a heavier concentration on selected categories of goods.

[4] Additional reasons for including the dummy are discussed below.

percent of the intercountry variation in export instability.

Although the problem of export instability has relevance mainly for *LDC*'s, I have included developed countries (*DC*s) as well in the analysis. One reason for this is to provide a wider range of variation in the explanatory variables. Second, including both *DC*s and *LDC*s in the sample permits one to test whether there is a difference in instability between the two groups of countries that cannot be explained by the other explanatory variables.

Ideally one would wish to determine, for any one country, the effect of changes in specific instrument variables. Because of data limitations, this is infeasible.[5] However, one can obtain information about the factors responsible for instability in a cross-country context. A number of studies have used cross-section data to relate intercountry differences in export instability to several structural variables.[6] A weakness in these studies is the absence of an explicit theoretical model explaining export instability and thereby justifying the particular assortment of explanatory variables chosen. The lack of a model makes the findings difficult to interpret. The present study attempts to improve upon the earlier work by developing an explicit model upon which the statistical estimation is based. This model relates instability in a country's total export receipts to fluctuations in individual exports, which in turn are explained in terms of the structural variables that appear in the regression analysis.

I. *Definition and Measurement of Instability*

In constructing an export instability index, it is necessary first to eliminate the trend. Otherwise, a country whose exports are growing rapidly—even at a constant rate—will score high on the instability scale. Measuring instability around the trend separates growth over the period as a whole from year-to-year deviations from the growth path. The latter constitute instability as the term is used here.

The type of trend fitted to the data influences the measures of instability obtained. There is some theoretical justification for using an exponential trend. Countries tend to plan in terms of the growth rate, not in terms of absolute increments. It is thus relevant to explain deviations from an exponential growth path. Also, for the time period used, an exponential trend provides a better fit to the data than a linear trend for most countries in the sample. Thus, for both theoretical and empirical reasons, I decided in favor of the exponential.

The instability index, denoted I, is defined as the standard deviation of the residuals from the trend. Letting X_{kt} = the value (in current dollars) of merchandise exports of country k in year t, I used ordinary least squares to estimate the following:

$$\log(X_{kt}) = a_k + b_k t + u_{kt} \qquad (k = 1, \ldots, K) \tag{1}$$

Then I_k = the standard deviation of the

[5] The export instability index, to be meaningful, must be based on a time-series covering at least a dozen or so years. To be able to obtain intertemporal comparisons for a country would require a period of some 30 years or more. Time-series data (of the sort required) for a period of this length is not available for most countries, and particularly for *LDC*s.

[6] In studies by Joseph Coppock, Alasdair MacBean, and Massell, samples ranging in size from 35 to 80 countries were used with an index of export instability calculated for 1948–59 (or in some cases 1946–59). Instability was found to be positively correlated with the degree of commodity concentration of exports and with the proportion of export receipts derived from primary products, and negatively correlated with per capita income and with the concentration of exports by geographic area of destination. The instability index and the explanatory variables were defined in a variety of ways, making it difficult to obtain a precise comparison of the different sets of results.

observed u_{kt}. Equation (1) was estimated and I_k calculated for each of 55 countries, using data for the period 1950–66.

II. *Explanation of Instability*

Consider that the (assumed exponential) trend in exports can be approximated by a linear function. Then the trend-corrected coefficient of variation of exports, denoted μ, will approximate the instability index, I.[7] In this section, I shall obtain an expression for μ in terms of the trend-corrected coefficients of variation of the individual export items.

The value of country k's export earnings from product i in year t, X_{kti}, can be written as a linear function of time (suppressing the k subscript),[8]

$$X_{ti} = \alpha_i + \gamma_i t + x_{ti} \tag{2}$$

where t = time (in years) and the x_{ti} = deviations from the linear trend. Let

$$X_t = \sum_i X_{ti}, \tag{3}$$

and

$$x_t = \sum_i x_{ti}, \tag{4}$$

where X_t = total export earnings in year t and x_t = total deviations from the linear trend. From (4) we can write

$$\sigma^2 = \sum \sigma_{ii} + \sum_i \sum_{j \neq i} \sigma_{ij} \tag{5}$$

where σ^2 and σ_{ii} denote, respectively, the variances of x_t and x_{ti} and where σ_{ij} = the covariance of x_{ti} and x_{tj}. Letting

$$\overline{X} \equiv \frac{1}{T} \sum_t X_t \quad \text{and} \quad \overline{X}_i = \frac{1}{T} \sum_t X_{ti},$$

we can write

$$\overline{X} = \sum_i \overline{X}_i \tag{6}$$

Equation (5) can be rewritten[9]

$$\mu = \left[\sum \lambda_i^2 \mu_i^2 + \sum \sum \lambda_i \lambda_j \mu_i \mu_j \rho_{ij} \right]^{1/2} \tag{7}$$

where

$$\mu = \frac{\sigma}{\overline{X}} \tag{8}$$

$$\mu_i = \frac{\sigma_i}{\overline{X}_i} \tag{9}$$

$$\lambda_i = \frac{\overline{X}_i}{\overline{X}} \tag{10}$$

ρ_{ij} = the coefficient of correlation between x_i and x_j. In (7), μ is the trend-corrected coefficient of variation of total exports which is an approximation to I, the index of instability. From (7), we see that μ is a function of the trend-corrected coefficients of variation of the individual export items, the composition of exports, as given by the set of λ_i, and the correlation coefficients, ρ_{ij}.

To explain intercountry differences in μ, we wish to focus on differences in the μ_i, λ_i, and ρ_{ij}. We begin first with a brief discussion of the determinants of X_{ti}. Consider Figure 1, where S, D_H, and D_T denote, respectively, the home supply, home demand, and total (home plus foreign) demand curves for an export

[7] The trend-corrected coefficient of variation is defined in equation (7). It is the standard deviation of the residuals from a linear trend of $X_t/\overline{X}$, where $\overline{X}$ is the arithmetic mean of the X_t.

[8] Strictly speaking, having found that the trend in exports is closer to an exponential than to a linear function, we should not write (2) as a linear relationship. However it seems acceptable for the heuristic argument of the present section to assume that the trend can be approximated by a linear function. We calculated μ and found it to be quite close in value to I for most countries. However, the linear trend gives a poor fit in several cases.

[9] From (5), we can write

$$\frac{\sigma^2}{(\overline{X})^2} = \sum \left(\frac{\sigma_i}{\overline{X}_i}\right)^2 \left(\frac{\overline{X}_i}{\overline{X}}\right)^2 + \sum \sum \rho_{ij} \left(\frac{\sigma_i}{\overline{X}_i}\right) \left(\frac{\sigma_j}{\overline{X}_j}\right) \left(\frac{\overline{X}_i}{\overline{X}}\right) \left(\frac{\overline{X}_j}{\overline{X}}\right)$$

from which (7) follows easily.

good. The equilibrium price is OP; OB is produced, OA sold domestically, and AB exported. (We assume no imports.) The value of exports of the good is given by the area of rectangle $AMNB$. This value can change as a result of shifts in any of the three curves. The effect on export earnings of a shift in one of the curves depends on the magnitude of the shift and on the elasticities of the curves. Thus the value of μ_i for any commodity depends on the relative instability and on the elasticities of the curves in Figure 1.

Supply Instability

Fluctuations in supply are more severe for some goods than for others. Some foods and agricultural raw materials experience especially great short-run supply instability. In the case of annual crops, supply fluctuations may arise from a cobweb effect. For both annuals and perennials, pests, plant disease, and weather variability often bring about sharp changes in supply. Countries whose exports are heavily weighted with goods with unstable supply curves will tend to experience greater instability in the value of exports.

The effect on a country's export receipts of a given degree of supply instability depends on the elasticity of foreign demand. The greater the departure of this elasticity from unity, the more pronounced is the effect on export receipts of a shift in supply. The elasticity of demand facing an individual country is influenced in part by the elasticity of the total demand curve for the product; but a much more important factor is the country's share of the world market. A country whose exports form only an insignificant part of world trade in a commodity can view the demand curve as virtually horizontal; whereas a country whose exports form a large part of world production will be faced with a demand curve with an elasticity closer to that of the total demand curve. In general, the smaller the country's share of the market, the more elastic the relevant demand curve will tend to be. If we assume that a country is faced with a less-than-unitary elastic demand curve only in rare circumstances, it follows that the departure of the demand elasticity from unity increases as the country's share of the world market declines. As a result, supply fluctuations can be expected to generate greater fluctuations in export receipts from a product the smaller the country's share of the world market for the product.

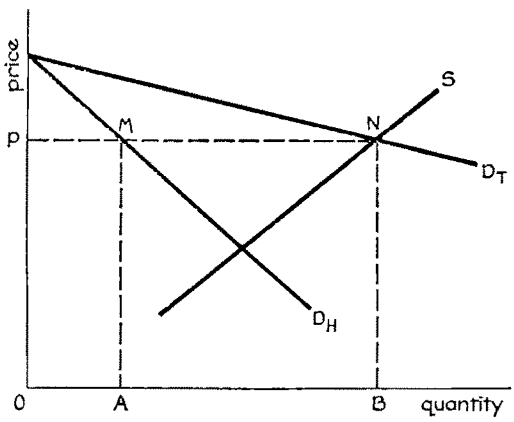


FIGURE 1

Demand Instability

For many products, shifts in foreign demand form the major source of fluctuations in export receipts. An individual country is confronted by a net foreign demand curve, based on aggregate world demand and supply. A shift in either the aggregate demand or the aggregate supply curve will alter the position of the net demand curve. Short-run shifts in aggregate supply result from the factors discussed above. Short-run shifts in demand may arise from changes in prices of competing goods or from cyclical changes in incomes (and thus in expenditures).

Some products are intrinsically more volatile than others. To a large degree this is due to demand-related factors. To the extent that shifts in demand arise from changes in purchasing power, these shifts will have a larger impact on an item the higher its short-run income elasticity. Foods, with a relatively low income elasticity, will tend to be less affected than most other goods. On the other hand, the demand for raw materials is often considered to be relatively unstable. Thus, countries that are heavily dependent on raw materials as a source of export receipts may tend to experience an above-average degree of export instability, whereas food-exporting countries may experience greater stability.

For a given foreign demand instability, and given elasticities, the coefficient of variation of export receipts is an increasing function of the proportion of domestic output consumed domestically (see Figure 1).[10] This follows because the larger domestic consumption is relative to exports, the greater will be the *relative* effect on the value of exports of a given absolute change in value (and thus of a given shift in demand).

The discussion thus far relates to foreign demand. However, the domestic demand curve may also be subject to short-run instability, possibly associated with inappropriate monetary and/or fiscal policy. Instability in domestic demand for a product (unless fortuitously offset by corresponding shifts in supply) tends to produce instability in the net export supply curve—i.e., the curve relating the net quantity offered abroad to the price. Fluctuations in the domestic demand curve therefore produce effects much like those of supply instability, discussed above.

[10] Although the values of the demand and supply elasticities influence the effect of a shift in demand, this influence is likely to be small. Regardless of the elasticities, the price and quantity effects reinforce each other.

Interrelationships Among Products

The instability of total export receipts depends not only on the instability of individual items but also on the correlation between receipts from different pairs of goods (see equation (7)).[11] Receipts from goods that are affected by similar market forces will tend to move together. Receipts from items that are dissimilar may fluctuate independently. In some cases, fluctuations in receipts from different goods may even offset each other. A country's total export receipts will tend to be more stable the more diversified are its exports—i.e., the larger the number of goods it exports, the more evenly its resources are spread over the different goods, and the more dissimilar these products are. A country with a large share of its exports derived from a single good or from several closely related goods will tend to experience greater instability than a country with a widely diversified export base.

Instability of exports may also be an increasing function of the geographic concentration of exports by country of destination. High geographic concentration is likely to imply greater dependence on economic conditions in one or a few countries. Fluctuations in demand in any recipient country will then have a more pronounced effect on receipts of the exporting country than if exports were more diversified among recipients. This is particularly likely to be the case if different markets are imperfect substitutes for one another as a result of trade preferences, bilateral trading arrangements, or other impediments to free multilateral trade.

III. *Definition and Measurement of the Explanatory Variables*

On the basis of the preceding discussion, export instability can be expected to be a

[11] This is discussed extensively by W. C. Brainard and R. N. Cooper.

function of: (a) the composition of exports; (b) the diversification of exports by commodity and by country of destination; (c) a country's share of world markets for the products that it exports; (d) the domestically consumed proportion of output of exported commodities. We shall devise indexes of each of these factors and use these indexes as explanatory variables in the regression equation.

Composition of Exports

Countries whose exports consist largely of goods with unstable demand and/or supply curves will tend to have greater export instability. In studies by Alasdair MacBean and by Massell, the proportion of the value of exports derived from primary products was used as an explanatory variable and found to be positively correlated with instability. Here the analysis is carried a step further and a distinction is made among three categories of goods: foods, raw materials, and manufactures. I hypothesize that the allocation of a country's export resources among the three categories influences the country's export instability. To test the hypothesis I have defined the following variables:

R_f = the proportion of total export earnings derived from foods, beverages, and tobacco (*SITC* groups 0, 1, 22, 4);

R_r = the proportion of export earnings derived from crude materials, petroleum, fuels, some fats, and base ores (*SITC* 2 (minus 22), 3, 67, 68).

For expository convenience R_f and R_r are termed, respectively, the food ratio and raw material ratio.

Export Diversification

My hypothesis is that export diversification tends to reduce the correlation between fluctuations in receipts from different items. The more widely a country's resources are spread among export products, and the more dissimilar these goods, the smaller will be the tendency for changes in receipts from different commodities to move in unison. Because of differences among covariances for pairs of commodities, it is difficult to measure effective diversification. However, as a first approximation, it is assumed that the earnings covariance will be lower between products in different three-digit *SITC* groups than between products in the same group. Then as an index of commodity concentration (or diversification) we use the Hirschman-Gini coefficient (Albert Hirschman 1945, 1964; Michael Michaely), denoted C, and written

$$C = \left(\sum y_i^2 \right)^{1/2} \tag{11}$$

where y_i = the proportion of total merchandise exports by value in *SITC* three-digit group i. The higher the value of C, the greater the concentration (the lower the diversification).

An intuitive way of looking at C was provided by Morris Adelman in a different context. Drawing on this work, we note that C is a function of the mean and variance of the value of exports in different three-digit groups. Of particular interest, C is the index of concentration that would result if a country's export receipts were divided evenly among $1/C^2$ different commodities. However C is an unambiguous measure of commodity concentration only if the product classification is meaningful—ideally, if $\rho_{ij}=1$ for any pair of goods in the same three-digit group and $\rho_{ij}=m<1$ for any pair of goods not in the same group. C was used as a measure of concentration by Massell and by MacBean and found to be weakly positively correlated with export instability. However C was highly multicollinear with the primary product ratio.

The measurement of geographic concentration is similar to that of commodity concentration. Letting G = the index of

geographic concentration, we write (Hirschman 1945; Michaely)

$$G = \left(\sum z_i^2 \right)^{1/2} \tag{12}$$

where z_i = the proportion of total exports shipped to country i. Massell and MacBean each found geographic concentration to be negatively related to export instability, contrary to the hypothesis above. However the relationship in both studies was statistically insignificant.

Country's Share of World Markets

It was argued that export instability is a decreasing function of a country's share of the world market in the goods it exports.[12] A country's importance as a supplier of world markets depends on the size of its export sector, the extent to which its exports are concentrated on a few items, and the size of the world market for these items. As an index (denoted Z_j) of the degree to which country j's exports tend to be large in world markets, we write

$$Z_j = \sum_i \lambda_{ij}\delta_{ij} \tag{13}$$

where

λ_{ij} = country j's share of commodity i in world trade and

δ_{ij} = commodity i as a proportion of country j's exports.

Then the product

$\lambda_{ij}\delta_{ij}$ = country j's share of commodity i weighted by the relative importance of i in j's exports.

Calculation of Z_j would be laborious. However one can obtain an approximate value of Z_j by considering the four most important exports (by value) for each country, calculating λ_{ij} and δ_{ij} for these items only, and then substituting these values in (13). This procedure was adopted here, using the *SITC* three-digit classification.

[12] Because of the assumption that the demand elasticity exceeds unity.

Domestic-Consumption Ratio

The earlier analysis suggests that instability is an increasing function of the extent to which exportable goods are consumed domestically. To obtain an index of the domestic consumption of exportables, I considered using the proportion of export receipts consisting of items over half of which were consumed at home. However, calculation of the domestic demand for raw materials and manufactures poses serious conceptual and measurement problems. It is more straightforward to obtain a measure of the home consumption of food items. Foods were divided into two groups according to whether the ratio of domestic consumption to total production was or was not greater than 50 percent. The proportion of the value of exports derived from foods in the greater-than-50 percent group is termed the domestic-consumption ratio, denoted D. It is hypothesized that D is positively associated with export instability.

Other Explanatory Variables

The six variables (C, G, R_f, R_r, Z, and D) discussed above are the ones suggested by the model. In this section, I discuss three additional variables that are included in the analysis: per capita income, a dummy variable to distinguish developed from less-developed countries, and size of the export sector.

Per capita income, denoted Y, is a rough measure of the level of economic sophistication of a country, which in turn is an important determinant of the types of goods produced. A more sophisticated country tends to produce—and export—products that are skill- and research-intensive. It is plausible to argue that the demand for such goods is relatively stable. A second point is that a high per capita income is usually associated with greater flexibility. A more flexible country is better able to shift resources among products in

response to (or in anticipation of) changes in demand conditions, and thereby to reduce the impact of sudden changes in demand. MacBean explored the relationship between export instability and per capita income and found a weak negative correlation. However, one should bear in mind that per capita income is highly correlated with several of the other explanatory variables, such as commodity concentration and the food ratio. While a simple correlation between I and Y may be negative, the question asked here is whether the relation is negative *net* of the other explanatory variables: i.e., how much of the relationship is explained by the variables discussed above?

MacBean's study noted that *LDC*s had a mean export instability 31 percent higher than *DC*s for the period 1948–59. In a recent paper Guy Erb and S. Schiavo-Campo found instability to be 117 percent higher in *LDC*s than in *DC*s for the period 1954–66. In the present study, for 1950–66, *LDC*s have 50 percent greater instability than *DC*s.[13] One would expect the difference in export instability between the two groups of countries to be largely attributable to differences in the explanatory variables discussed above. It is nevertheless of interest to see whether there is a difference in instability between *DC*s and *LDC*s over and above what can be explained by the other variables. Therefore I have added a dummy variable, M, to distinguish between the two groups of countries.

The final variable included in the analysis is the value of exports. Initially I argued that, because Z in equation (13) would be laborious to calculate, one could use the size of the export sector as a crude approximation to Z. Given the commodities exported, and given the concentration of exports, a country's share of world markets is likely to be an increasing function of the size of its export sector. This assumes that differences in the size of world markets for different goods will wash out—that is, that both large and small countries tend to export products whose world markets are comparable in size. This last assumption probably does not hold, but the value of exports, X, turned out to be significant at the 5 percent level.

The referee suggested the procedure that was later adopted: to approximate Z on the basis of each country's four most important exports. Although there does not appear to be any further theoretical justification for leaving X in the model, its significance in the earlier regressions suggests that it may be an important factor. Therefore I decided to leave it in the model. This leaves a total of nine explanatory variables.

IV. *Specification of the Model*

The regression equation is written

$$\begin{aligned} I = {} & b_0 + b_1C + b_2G + b_3R_f \\ & + b_4R_r + b_5 \log (D) \\ & + b_6 \log (Z) + b_7 \log (Y) \\ & + b_8M + b_9 \log (X) + \epsilon \end{aligned} \tag{14}$$

where

I = the instability index,
C = commodity concentration index, defined by (11),
G = geographic concentration index, defined by (12),
R_f = the food ratio,
R_r = the raw materials ratio,
D = the domestic-consumption ratio,
Z = the export market share coefficient,
Y = per capita income, in current *U.S.* dollars,
$M = \begin{cases} 1 \text{ for } DC\text{s} \\ 0 \text{ for } LDC\text{s} \end{cases}$
X = value of merchandise exports, in current *U.S.* dollars, and
ϵ = the disturbance.

[13] See Table 1. Note that a different instability index is used here from that used by Erb and Schiavo-Campo and by MacBean.

TABLE 1—SUMMARY STATISTICS OF SAMPLE DATA

	I	C	G	Y	Z	X	D	R_f	R_r
55 countries	.126	.408	.389	348[a]	.089	533[b]	.146	.413	.350
19 *DC*s	.095	.241	.337	1,213[a]	.069	1,840[b]	.093	.212	.301
39 *LDC*s	.142	.491	.421	185[a]	.099	266[b]	.182	.521	.378
Ratio *LDC*s to *DC*s	1.495	2.037	1.249	.153	1.435	.145	1.957	2.458	1.256

[a] *U.S.* dollars.
[b] Millions of *U.S.* dollars.

The variables C, G, R_f, and R_r all enter the regression linearly; the empirical distributions of these variables are sufficiently close to normal to make this quite acceptable. However, D, Y, Z, and X approximate more closely a lognormal distribution, so *logs* were used for these four variables.

V. *Data*

The sample contains 36 *LDC*s and 19 *DC*s chosen on the basis of the availability of reasonably reliable data. For C, G, R_f, R_r, Z, and D, data for 1960 or the closest available year were used.[14] For Y and X, we used the mean of the annual values for 1958–62. In determining M, we distinguish between *DC*s and *LDC*s principally on the basis of per capita income, with the dividing line set at \$600 dollars. However, Japan (with a per capita income of \$420) was considered a *DC*, and Venezuela (with a per capita income of \$978) was considered an *LDC*.

Table 1 presents the means of all variables used in the analysis for the *LDC*s, *DC*s and the full sample. As already noted, I is 50 percent higher for the *LDC*s. Y and X are naturally much lower, whereas C, Z, G, D, R_f, and R_r are all higher for the *LDC*s. It is interesting to note that $1/C^2$ equals 16 for *DC*s and 4 for *LDC*s.[15]

[14] See the Appendix for the data sources and list of countries.

[15] The interpretation of $1/C^2$ was discussed above.

VI. *Empirical Results*

Equation (14) was estimated using ordinary least squares. The residuals were inspected and found acceptable. The results are shown on the first line of Table 2. The R^2, corrected for degrees of freedom, is .414, which is significant at one percent using an F-test. C and R_f are significant at 5 percent but the remaining variables are insignificant at 10 percent. Four variables—G, Y, Z, and R_r—have estimated coefficients well under their standard errors; thus each individually reduces the corrected R^2. These four variables were deleted from the regression and the remaining coefficients reestimated.[16] The results appear on the second line of the table.

In line 2, R^2 is slightly higher (.457) and the five t-ratios are all larger. Both R_f and C are significant at 1 percent, and X at 5 percent. The estimated coefficients in line 2 are very similar to those in line 1, suggesting stability of the relationship with respect to the deletions. Reintroducing any of the four deleted regressors serves only to reduce R^2; the subset of five regressors whose coefficients appear in line 2 provides the best fit to the data.

On the basis of the t-ratios in line 2, the most highly significant variable is R_f. The high negative significance of R_f suggests that countries that derive a large percentage of their export earnings from food

[16] Each of these variables is collinear with several of the remaining variables.

TABLE 2—ESTIMATED REGRESSION COEFFICIENTS

R^2	F-ratio	Explanatory Variables*								
		C	D	X	R_f	M	Z	Y	R_r	G
.414[a]	5.23	.148[b] (2.37)	.014 (1.44)	−.010 (−1.62)	−.120[b] (−2.19)	−.025 (−1.04)	−.003 (−.35)	.003 (0.28)	−.014 (−0.23)	.014 (0.27)
.457[a]	10.09	.131[a] (3.03)	.012 (1.58)	−.010[b] (−2.29)	−.106[a] (−4.32)	−.019 (−1.20)				
.429[a]	9.10	.180[a] (4.21)	.014 (1.83)		−.101[a] (4.05)	−.024 (−1.43)	−.009 (−1.60)			

* See text for description. Figures in parentheses are t-ratios. Significance levels, using a two-tail t-test, are indicated as follows:
[a] 1 percent.
[b] 5 percent.

tend to experience less export instability than countries that are more heavily dependent on raw materials or manufactures.[17] This result is consistent with the notion that instability of a country's export receipts is to a large extent the result of year-to-year shifts in the demand for its exports. If instability results mainly from shifts in the supply curves, then one would expect agricultural products—foods and (to an extent) raw materials—to experience greater fluctuations than manufactures.[18] On the other hand, if instability is the result of fluctuations in the general level of demand, it will tend to affect most goods but will have greater impact on goods with high short-run income elasticities. As foods tend to be relatively income inelastic, they will experience less sharp fluctuations.

The coefficient of C is significantly positive at one percent,[19] supporting the hypothesized relationship between instability and commodity concentration. This finding is also consistent with the view that shifts in foreign demand have been a major of export instability. Diversification would be relatively ineffective in reducing fluctuations due to shifts in a country's supply curves.

The coefficient of X is significantly negative at 5 percent. I originally argued that the negative significance of X supports the hypothesis that countries with small export sectors tend to be confronted with more elastic foreign demand curves for their products and hence are more susceptible to supply-induced fluctuations. However, on a priori grounds, one would expect Z to be a better measure of export market share, and the insignificance of Z casts doubt on this argument. We return to this point below.

The coefficient of D is positive but (barely) insignificant at 10 percent. This provides weak support that instability tends to be greater for a country exporting goods that are mainly consumed domestically. This result, in turn, is consistent with the importance of shifts in foreign demand as a source of instability.

The dummy variable, M, has a negative estimated coefficient greater in absolute

[17] This relationship holds only net of the other explanatory variables; the simple correlation between I and R_f is −.03.

[18] It is possible also that countries with a high R_f have established preferential markets for their exports, tending to reduce instability of food exports.

[19] The significance of C is substantially higher than in the MacBean and Massell studies.

value than its standard error but statistically insignificant at 10 percent. Although this result is thus inconclusive, there is a slight suggestion that *LDC*s tend to experience greater instability than *DC*s, net of the other explanatory variables. As measured by the constant term and the dummy variable term in (13), mean instability is .120 for *DC*s and .139 for *LDC*s. Net of the explanatory variables, therefore, *LDC*s have experienced 16 percent greater instability than *DC*s. Without taking account of differences in the explanatory variables, the mean instability index for *DC*s is .095 and for *LDC*s is .142, 50 percent greater. Differences in C, R_f, D, and X therefore reduce the original discrepancy from 50 to 16 percent.

Geographic concentration was found to have a negative estimated coefficient insignificant at 5 percent in earlier studies. In the present study, G has a t-ratio of less than one in all regressions in which it was included, and thus appears to be unimportant as an explanatory variable. This suggests that world markets perform well despite the existence of trading blocs.

The statistical insignificance of R_r contrasts with the high significance of R_f. Export instability in the sample is apparently not influenced by the allocation of export resources between manufactures and raw materials. The insignificance of R_r and the negative significance of R_f suggest that there is no tendency for primary products to be more unstable than manufactures (net of the other variables).This finding is contrary to results obtained by MacBean and by Massell, who found the primary product ratio to have a significantly positive coefficient.

Per capita income also failed to have a t-ratio of one in any regression and thus appears to be unimportant in determining intercountry differences in export instability. As per capita income is highly correlated with the dummy variable ($r=.83$), a natural thought is that the standard error of the Y coefficient blew up due to multicollinearity. Thus we substituted Y for M and reran the regression. This lowered R^2 and left Y with a t-ratio under unity. Consequently the greater export instability of *LDC*s is apparently not a result of their lower per capita income.

The regression was reestimated substituting Z for X; the results appear in line 3 of Table 2. R^2 is slightly lower than in line 2 (.429 vs. .457) but three of the t-ratios are higher. R_f and C are still both significant at 1 percent, D is nearly significant at 5 percent, and Z nearly significant at 10 percent. The only coefficient that changes markedly is that associated with C, which is higher in line 3 than in line 2.

If both X and Z are included in the regression, Z has a t-ratio of only $-.49$ and C a t-ratio of only 2.92. The results thus suggest considerable multicollinearity between X and Z and, to a lesser extent, between X and C. This is not surprising. However, it is surprising that X provides a better fit than Z, considering that on a priori grounds, Z is a more obvious choice for inclusion in the model. Possibly the size of a country's export sector has an effect on the stability of the country's export earnings in ways not fully explained by the model.

VII. *Concluding Remarks*

The analysis presented here considers the contribution of several explanatory variables to intercountry differences in export instability. The results are influenced by the particular sample chosen and by the years upon which the instability index is based. The period used here contains the Korean War, which contributed greatly to the instability in export receipts. Use of a different period might produce different results. In this sense, one must be careful in making inferences.

Another qualification concerns the rel-

evance of cross-sectionally estimated parameters for answering questions that are implicitly of a time-series nature. A planner in an *LDC*, who wishes to reduce export instability, can form some general idea of the problem from our results but they do not provide all the relevant information for his particular country. Some variables that affect export instability are difficult to quantify. These variables differ among countries, making it hard to apply the results to the problem of a particular country.

In attempting to form policy judgements, one would want to know the cost of reducing instability. For example, the cost of diversifying a country's export sector depends on such fundamental matters as comparative advantage. If diversification involves shifting resources into substantially less productive uses, then the cost will be large.

The gains achievable from reduced instability also need more careful consideration. To what extent are fluctuations in export receipts transmitted to the economy?[20] What effect does this have on the rate of growth? What effect on the distribution of income? Without answering such questions it is difficult to reach meaningful policy judgements.

As noted earlier, the *LDC*s have high values of C and D and a low value of X, all of which help explain the greater export instability of *LDC*s. The one mitigating factor is the *LDC*s' dependence on food exports which, as we have seen, tend to be more stable (net of the other regressors). If a country is small and exports only a few items (presumably because it is small), it does well to concentrate on foods. Food producers also tend to be more highly concentrated; the coefficient of correlation between R_f and C is .43. If a food producer can diversify its exports it would apparently reduce its instability. Instability is reduced also by shifting into items that are primarily export goods. Of course, given a country's resources, it is difficult to expand food production so as to reduce C and D jointly. As most foods are consumed in some measure domestically, a reduction in D may require producing larger amounts of selected goods, thus tending to increase C. One way around this is to concentrate on beverage crops such as coffee and cocoa, and to diversify within this class of goods. Unfortunately, for most countries this is infeasible.

Finally, it is interesting to note the quantitative implications of individually altering each of the explanatory variables from the *LDC*s' mean value to the *DC*s' mean. This is shown in Table 3 based on the line 2 and line 3 sets of estimated coefficients. The variables with the largest impact are R_f and C. However the two variables tend to offset each other. *LDC*s tend to experience greater instability because of their greater product concentration but less instability as a result of their heavier dependence upon foods.

TABLE 3—EFFECT OF EXPLANATORY VARIABLES IN EXPLAINING DIFFERENCES IN INSTABILITY BETWEEN *LDC*s AND *DC*s

Variable	Mean Value		Percentage Change in I resulting from Change from *LDC* Mean to *DC* Mean	
	*LDC*s	*DC*s	Line 2 Coefficients	Line 3 Coefficients
R_f	.521	.212	+23	+22
C	.491	.241	−23	−32
D	.182	.093	− 6	− 7
X	.266	1,840	5	...
Z	.099	.069	...	+ 3

APPENDIX

A. *Countries Included in the Sample* (1) *LDC*s: Argentina, Bolivia, Brazil, Burma, Ceylon, Chile, Colombia, Costa Rica, Do-

[20] This is investigated in the forthcoming paper by Massell, Pearson, and Fitch.

minican Republic, Ecuador, El Salvador, Ethiopia, Greece, Guatemala, Honduras, India, Indonesia, Jordan, Mexico, Nicaragua, Nigeria, Pakistan, Paraguay, Peru, Philippines, Portugal, South Africa, Spain, Sudan, Taiwan, Thailand, Turkey, United Arab Republic, Uruguay, Venezuela, Yugoslavia; (2) *DCs*: Australia, Austria, Belgium-Luxembourg, Canada, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, United States.

B. *Data Sources*. Merchandise exports in current dollars, 1950–66 (for calculation of I): *IMF, Balance of Payments Yearbook*, vols. 5, 7, 13, 15, 19; data for Spain and Nigeria from *AID* files. Exports averaged 1958–62 for calculation of x.

Exports in 1960 by country of destination (for calculation of G); and exports in 1960 by *SITC* classification (for calculation of C, R_f, R_r, Z): United Nations, *Yearbook of International Trade Statistics*. (*SITC* for Chile, China, Ethiopia, Sudan: 1962: for Jordan: 1964; for Peru: 1961; for Bolivia, Paraguay, Uruguay, *SITC* classification approximated using available data.)

Population, 1960; and Gross National Product, 1960: *IMF, International Financial Statistics*, supplement to 1966–67 issues. *GNP* averaged 1958–62, except: El Salvador, Nigeria, Pakistan, *GDP*; India, *UAR*, National Income; Jordan, *GNP* 1959–62; Peru *GNP* 1958–60; Yugoslavia, Gross Material Product; *GNP* for Ethiopia, Indonesia, Surinam from *AID* files.

Domestic Consumption of export commodities: *IMF, International Financial Statistics; UN, Yearbook of International Trade Statistics;* Gill and Duffus, *Cocoa Market Report; USDA, World Coffee Crop.*

REFERENCES

M. A. Adelman, "Comment on the 'H' Concentration Measure as a Numbers-Equivalent," *Rev. Econ. Statist.*, Feb. 1969, *51*, 99–101.

W. C. Brainard and R. N. Cooper, "Uncertainty and Diversification in International Trade," *Food Res. Inst. Stud. 8*, No. 3, 1968, 257–85.

J. D. Coppock, *International Economic Instability*, New York 1962.

G. F. Erb and S. Schiavo-Campo, "Export Instability, Level of Development, and Economic Size of Less Developed Countries," *Bull. Oxford Univ. Inst. Econ. Statist.*, Nov. 1969; *31*, 263–83.

A. Hirschman, *National Power and the Structure of Foreign Trade*, Berkeley 1945.

———, "The Paternity of an Index," *Amer. Econ. Rev.*, Sept. 1964, *54*, 761–62.

A. T. MacBean, *Export Instability and Economic Development*, Cambridge, Mass. 1966.

B. F. Massell, "Export Concentration and Fluctuations in Export Earnings: A Cross-Section Analysis," *Amer. Econ. Rev.*, Mar. 1964, *54*, 47–63.

———, S. R. Pearson, and J. B. Fitch, "Foreign Exchange and Economic Development in Latin America," forthcoming.

M. Michaely, *Concentration in International Trade*, Amsterdam 1962.

Inflation, Unemployment, and Economic Welfare

By ROGER N. WAUD*

It is presently believed that anticipated inflation has an unambiguously bad effect upon "feasible welfare."[1] In this paper it will be shown that if money wages are not flexible, and hence positions of unemployment equilibrium are possible, then the expectation of inflation can have good effects on feasible welfare. It is then assumed that the three main goals of economic policy are the achievement of "full-liquidity" (a situation where the opportunity cost of holding money balances is nil), the maintenance of a real rate of interest deemed to give rise to an "optimum" division of output between investment and consumption goods (or an optimum rate of economic growth), and the reduction of unemployment or the achievement of full-employment.[2] Given these assumptions, it will be shown that while the expectation of inflation has a negative effect on attempts to achieve full-liquidity, it has an unambiguously positive effect on attempts to achieve full-employment and *may* have a positive effect on attempts to maintain a real rate of interest consistent with optimum growth. The upshot of these opposite effects on economic welfare resulting from inflation is to strengthen the case for the payment of an appropriate rate of interest on money, because as will be shown, the ability of the monetary-fiscal authority to achieve simultaneously both the employment and growth goals is greatly enhanced if interest is paid on money at a rate sufficient to achieve full-liquidity.

I. *Anticipated Inflation and Unemployment Equilibrium*

In the model employed, prices are taken to be flexible, while it is assumed that money wages do not fall when there is an excess supply of labor.[3] Assuming that the capital stock is fixed,[4] the usual equilibrium in the labor market requires that

$$\frac{\overline{W}}{p} = y_N(N) \tag{1}$$

where $\overline{W}$ is the fixed money wage, p is an index of the price level applicable to real output y, N is the volume of employment (in equivalent fulltime workers), and $y_N(N)$ is the first derivative of the production function $y(N)$ with respect to N.

We suppose that total wealth consists of only two assets, M, noninterest-bearing "money," and, a, equity "shares" (claims upon capital goods). The private sector is assumed to hold a proportion λ of

* Associate professor of economics, University of North Carolina, Chapel Hill.

[1] This view has been formally expressed by Milton Friedman (1953, 1969), Martin Bailey, and Edmund Phelps.

[2] Phelps has introduced monetary and fiscal controls into Lloyd Metzler's flexible wage-price model in order to relate monetary and fiscal policy to two of these goals—achievement of full-liquidity and maintenance of the optimum real rate of interest.

[3] This money wage rigidity may be attributed to any number of factors. Some of the most commonly cited are: minimum wage laws and other such statutory provisions; powerful trade unions; unwillingness of employers to reduce money wages due to a desire to retain experienced and skilled employees; unwillingness of unemployed workers to accept reduced money wages even though they would be willing to work at lower real wages brought about by a rise in prices. An interesting discussion of this issue is given by Axel Leijonhufvud (pp. 81–111).

[4] It is assumed that the capital stock changes by discrete amounts in every period.

these shares while the central bank or monetary authority holds the remaining proportion $1-\lambda$; a change in the relative proportions held may be achieved only through open-market operations (i.e., a change in λ). The real value of equity shares is the capitalized value of all business profits,

$$a = \frac{cy}{r} \tag{2}$$

where c is the proportion of real output consisting of business profits and r is the real rate of interest.[5] Real private wealth is then

$$w = \lambda a + m \tag{3}$$

where m is real money balances.[6]

The demand for real money holdings is assumed to be a function of the nominal rate of interest i (i.e., the prospective nominal yield on shares which may be viewed as the cost of holding money) and the level of aggregate real output y.[7] Presumably, the higher the cost of holding money (i.e., the higher the nominal yield on shares), the smaller is the amount of money held relative to transactions or real output y. The demand function for real money balances is then

$$m = \frac{M}{p} = yk(i) \tag{4}$$

where $k_i<0$, $\partial m/\partial y>0$, and M is the nominal stock of money.

If we take the nominal supply of money to be determined by the fiscal and monetary authorities, then $M=\overline{M}$ and

$$m = \frac{\overline{M}}{p} \tag{5}$$

The market in which money and shares are traded will be in equilibrium when the price level, the prospective nominal yield on shares, and the level of production or real output y equate the demand and supply of money, i.e., when equations (4) and (5) are simultaneously satisfied.

The amount of output I (measured in real terms) invested in any period is assumed to be inversely related to the real rate of interest; i.e., the prospective real rate of return on shares; hence $I=I(r)$, $I_r<0$. In any period, I is equal to private saving (unconsumed production), S, plus public saving (the difference between government expenditures G and tax receipts R) which is simply the government budgetary surplus or net taxes T. That is

$$I = S + R - G = S + T$$

Taxes R are taken to be lump sum.[8]

It is assumed that private (real) saving depends on the real rate of interest (the real prospective yield on shares), real private wealth, real income, and net taxes; thus $S=S(r, w, y, T)$. Furthermore, we assume that an increase of wealth will decrease desired private saving, $S_w<0$; that a rise of taxes reduces saving but by less than the amount of the tax, $-1<S_T<0$; and that an increase of the real return on shares will increase saving, $S_r>0$. Assuming net taxes constant $(T=\overline{T})$, equilibrium in the goods market requires that the real rate of interest, real wealth, and

[5] The assumption that profits are a constant share of income in a model of unemployment equilibirum is somewhat implausible. It doesn't affect any major conclusion however and is therefore maintained in the interest of simplicity.

[6] It is assumed that the real value of privately held shares, λa, in any period equals the real value of the economy's capital after the investment of the current period.

[7] We assume that the aggregate volume of planned transactions depends mainly on the volume of current production and that the primary motive for holding money is the transactions demand.

[8] Although the system will react somewhat differently depending on whether a change in T results from a change in G or R, the overall conclusions are not affected by this difference. Looking at the effects on the system resulting from changes in T, as we will do subsequently, is equivalent to analyzing the effects of changes in taxes or R for a given level of G.

real income be such as to equate the corresponding level of investment to the corresponding level of saving:

$$(6) \qquad I(r) = S(r, w, y, \overline{T}) + \overline{T}$$

The above model is presented diagrammatically in Figure 1. The EE schedule is the locus of real interest rates and nominal income levels py which give equilibrium in the goods market; it is defined by equations (1), (2), (3), (5), and (6).[9] When deriving the locus of real interest rates and nominal income levels consistent with equilibrium in the money market, it must be recognized that if the price level (of commodities and equity shares) is rising, or is expected to, then there is the additional opportunity cost of holding money which is equal to the rate of depreciation of its purchasing power with respect to commodities and equity shares (i.e., the rate of increase of the price level). Conversely, if the price level is falling, or is expected to, then the cost of holding money balances is reduced (or subsidized) by the rate of appreciation of its purchasing power (i.e., the rate of decrease of the price level). If ρ denotes the expected rate of change of the price level, then the relationship between ρ, the real rate of interest r, and the cost of holding money or nominal rate of interest i is

$$r = i - \rho$$

[9] In order to see that the EE schedule is negatively sloped, suppose the goods market to be in an equilibrium characterized by point a on the EE schedule (Figure 1). Now suppose the real rate of interest is dropped to a lower level. The level of investment will now be larger since $I_r<0$. On the other hand, the fall in r will cause saving to fall since $S_r>0$, $S_w<0$ and w rises according to (2) and (3). In order to restore equality in (6), y will have to be larger since $S_y>0$. Assuming the second derivative of the production function with respect to N is negative, a rise in y will also mean a rise in p according to (1), which will also tend to increase saving according to (5), (3) and the assumption that $S_w<0$. Hence, for a lower r, equilibrium in the goods market requires a larger y and p—hence a higher nominal income, py.

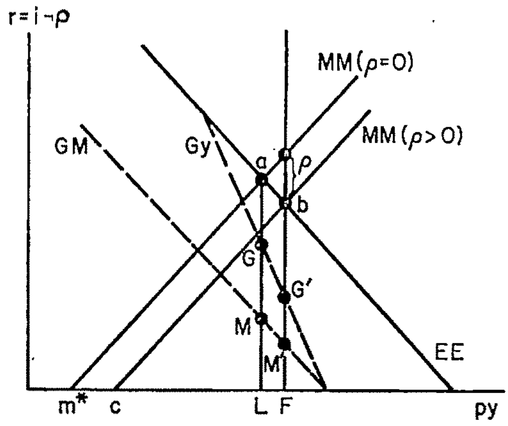


FIGURE 1

Assuming ρ equals zero, the locus of real rates of interest and nominal income levels giving equilibrium in the money market is given by the $MM(\rho=0)$ schedule in Figure 1; this schedule is defined by equation (4) in conjunction with the supply condition that $M=\overline{M}$.[10]

Suppose that the system is in a position of less than full-employment equilibrium determined by the intersection of EE and $MM(\rho=0)$ at a in Figure 1. The underemployment equilibrium nominal income level is L while the nominal income level necessary for full-employment is F. The anticipated rate of price change ρ is zero. Suppose now that an exogenously anticipated rate of inflation develops, i.e., ρ becomes positive. At any real rate of interest, r, the opportunity cost of holding money balances increases by the amount ρ and $k(i)$ becomes smaller since $k_i<0$. Hence at any real rate of interest, the level of py consistent with equilibrium in the money market is now larger; the MM schedule is therefore shifted rightwards. If the rise in the exogenously anticipated rate of infla-

[10] It is readily shown that the MM schedule is positively sloped. Assume the money supply to be fixed at $M=\overline{M}$ and, for simplicity, that $\rho=0$ so that $i=r$. If r falls, simultaneous satisfaction of (4) and (5), i.e., maintenance of equilbrium in the money market, requires a lower level of py, given that $k_i<0$.

tion ρ is large enough, it is possible that the MM schedule could shift to the position $MM(\rho>0)$ where it intersects EE at b (Figure 1), thus lowering the real rate of interest and raising nominal income to the full-employment level F. In any event, the onset of anticipated inflation will move the system towards full-employment.

The realization of a higher output level, the associated reduction in unemployment, and the concomitantly higher level of resource utilization, all resulting from the expectation of inflation, constitute an increase in economic welfare in and of themselves. We thus arrive at an unfamiliar, and heretofore oft-denied, proposition: anticipated inflation is not unambiguously bad.

In addition to the full-employment objective, suppose that policy makers also desire to achieve full-liquidity and a real rate of interest deemed to lead to an optimum division of total output between investment and consumption goods. For the moment, ignore the latter objective and consider the possibility of anticipated inflation simultaneously leading to the achievement of the first two.

If full-liquidity is said to exist whenever the desire to economize on money is nil, such a situation can occur only when the opportunity cost of holding money is zero.[11] This means that whatever the real output level, the demand for real money balances is given by equation (4) with i equal zero or

$$m = k\cdot y \tag{4'}$$

Equation (4′) in conjunction with (5) determines the level of nominal income consistent with full-liquidity. In Figure 1 this is given by m^*, the point where $MM(\rho=0)$ meets the py axis.

Note that the equilibrium nominal income level L associated with a zero rate of anticipated inflation is closer to m^* than is F, the equilibrium nominal income level associated with a non-zero rate of anticipated inflation. Hence while anticipated inflation has the desirable consequence of moving the system toward a full-employment equilibrium position, it has the undesirable effect of increasing the short fall from a full-liquidity position. Anticipated inflation therefore contributes to economic welfare in the former sense while detracting from it in the latter.[12]

In addition to the full-employment and full-liquidity objectives, consider now the possibility of achieving a real rate of interest which gives rise to a division of total output between investment and consumption goods which is deemed optimum. Presumably, this objective is meaningful only if the full-employment output level has been achieved. Otherwise the division of any total output level between investment and consumption goods can be improved by simply increasing output to the full-employment level since more of both goods will then be available. In the absence of monetary and/or fiscal intervention, the expectation of inflation will lead to the achievement of both full-employment and the optimum division of output only if the real interest rate level consistent with full-employment is coincidentally the same as that giving rise to the optimum division of the full-employment total output level.

In general then, while inflationary expectations will move the system toward

[11] The term "full-liquidity" was first coined by Phelps.

[12] It has been suggested by Friedman and Phelps that an induced deflation would be one possible way of eliminating, or at least alleviating, the opportunity cost of holding money. An anticipated rate of deflation, i.e., $\rho<0$, would cause the MM schedule to intersect the EE schedule of Figure 1 at a point to the left of a, thus giving an equilibrium nominal income level to the left of L and closer to m^*. However it is obvious that when positions of less-than full-employment equilibrium are possible, induced deflation, while moving the system closer to the desirable full-liquidity position m^*, has the undesirable consequence of simultaneously moving the system further away from the full-employment equilibrium nominal income level F.

the full-employment objective, they move it away from the full-liquidity objective and, even if they lead to a real interest rate level consistent with full-employment, there is no reason to assume it is the level consistent with an optimum division of that output between investment and consumption.

II. *Monetary and Fiscal Policy and the Welfare Objectives*

Consider now the feasibility of achieving the welfare objectives by use of monetary and fiscal policy; first, in the absence of inflationary expectations, and then when they are present.

In the absence of inflationary expectations, it is possible to achieve full-liquidity by use of fiscal and monetary policy only in the unlikely event that the *EE* and *MM* schedules are made to intersect at a zero real rate of interest, i.e., at the point m^* in Figure 1. In order to assess the possibility of achieving the other two welfare objectives, it is necessary to totally differentiate the whole system with respect to the parameters under the control of the monetary-fiscal authorities; namely, the proportion λ of total outstanding equity shares held by the private sector, the nominal money stock M, and the government budgetary surplus, or net tax level T. This is done in the Appendix; the results relevant to the question under consideration here are shown in Table 1.

TABLE 1—THE EFFECTS ON REAL OUTPUT, y, AND THE REAL INTEREST RATE, r, OF OPEN MARKET OPERATIONS AND CHANGES IN THE NET TAX LEVEL, T[a]

	dT	$d\lambda$	dM
dr	$\frac{dr}{dT} < 0$	$\frac{dr}{d\lambda} > 0$	$\frac{dr}{dM} \lesseqgtr 0$[b]
dy	$\frac{dy}{dT} < 0$	$\frac{dy}{d\lambda} > 0$	$\frac{dy}{dM} > 0$

[a] These results are predicated on the reasonable assumption that

$$-Mg_y + k(r)\overline{W} > -r\overline{W}k_r$$

as discussed in the Appendix.

[b] In the Appendix, it is shown that

$$\frac{dr}{dm} \gtrless 0$$

when

$$-S_w\left[\frac{\lambda c}{r} + k(r)\right] - S_y \gtrless 0$$

Since λ is the proportion of the real stock of securities in the private sector, an open market purchase will reduce λ while increasing M, conversely for an open market sale. The ultimate consequence of an open market operation is thus determined by considering the effect on the system of the change in both λ and M. Table 1 indicates that open market operations affect both y and r; y ambiguously and r possibly so. Similarly, a change in T, the level of the government budgetary surplus or deficit, will cause both y and r to change, in the opposite direction to the change in T.[13] These considerations strongly suggest that there is no guarantee that alternate adjustments in the fiscal variable T and the monetary variables M and λ would ever lead to the simultaneous attainment of the full-employment income level and the optimum real interest rate. Simultaneous achievement of all three welfare objectives would appear inconceivable a fortiori.

In the presence of exogenous inflationary expectations the problem of simultaneously achieving the three welfare objectives is the same as that encountered in their absence because the changes in y and r consequent upon changes in the monetary and fiscal parameters λ, M, and T are the same as those shown in Table 1. However, these results are predicated on the

[13] Note that any change in the budgetary surplus T implies a change in the rate of change of the money supply, not a change in the money supply. I am indebted to Arthur Laffer for drawing my attention to this point.

assumption that the expectation of inflation (or deflation) is exogenously given. It is more realistic to assume that inflationary (or deflationary) expectations are *induced* by *actual* rates of inflation (or deflation).

Proceeding on this assumption it should be recognized that since there are just two assets in the model, money and shares, any budget deficit (or surplus) T must be "financed" by a continual rate of change of the money supply which will be positive (negative) when there is a budget deficit (surplus). Any change in T therefore implies a change in the rate of change of the money supply, *not* a change in the money supply. Positive savings-investment (in excess of replacement needs) implies a growing stock of capital goods and hence a positive rate of growth of output; a change in the savings-investment level implies a change in the rate of growth of output. Since these growth effects are usually dismissed as of secondary importance in the time interval relevant to the short-run equilibrium under consideration, the equilibrium is not complete, as is generally admitted in this type of static analysis. Ignoring the effects of the rate of change of the money supply implied by a non-zero value of T is analogous to ignoring the growth effects implied by a positive savings-investment relation. However it can be shown that ignoring these so-called "secondary" effects leads to a fallacy in the interpretation of the static type of macroeconomic equilibrium under consideration here.[14] Moreover, it will be seen that recognition of these effects is crucial to understanding the role of induced inflation (or deflation) in this analysis.

Now reconsider the equilibrium discussed in Figure 1. There it was shown that if the system was in an underemployment equilibrium with income level L, the onset of inflationary expectations would cause the MM schedule to shift rightwards to a position such as $MM(\rho>0)$ giving a higher income and employment level,[15] though, of course, not necessarily the full-employment income level F depicted in Figure 1. However such an *exogenous* onset of anticipated inflation (or deflation) is a much less likely development than the *induced* onset of anticipated inflation (or deflation) brought about by an *actual*, or observable, rate of price inflation (or deflation). The induced expectation of price change at some rate ρ is more relevant when explicit recognition is given to the actual rate of change of the nominal money stock, implied by a non-zero T, and the actual rate of growth of output implied by a positive savings-investment level. If the money supply is increasing at a higher rate than the rate of growth of total output, prices will be rising and inflationary expectations will be induced; deflationary expectations will be induced if the reverse is true.

In Figure 1 the vertical distance (measured on the vertical axis) between the EE and GM schedules measures the percentage rate of growth of the nominal money supply resulting from a budget deficit.[16] The vertical distance between the EE and the Gy schedules measures the percentage rate of growth of real output. This distance gets larger as the real rate of interest falls reflecting the presumption that the lower the real rate r, the higher the level of real investment and hence the higher the rate of growth of capacity, and therefore the larger the rate of growth of real output. It is now obvious that the "equilibrium" given by the intersection of $MM(\rho=0)$ and EE at a is not complete. For at this point the percentage rate of growth of the nominal money stock (equal

[14] See Robert Mundell.

[15] This will always be true so long as the elasticity of demand for money balances with respect to the nominal interest rate i (hence with respect to ρ) is not zero, i.e., so long as $k_i<0$.

[16] GM is parallel to EE reflecting the fact that the tax is lump sum.

Ma) exceeds that of real output (equal Ga) by the amount MG which is therefore the *actual* percentage rate of rise of prices; since at a the actual rate of inflation MG exceeds the anticipated rate ρ (equal zero), expectations are not consistent with realizations—a necessary condition for equilibrium.[17]

If ρ is to equal the actual rate of inflation, the MM schedule in Figure 1 must lie to the right of $MM(\rho=0)$. To the right of a the actual rate of inflation associated with the intersection of MM and EE will be lower because the gap between GM and Gy narrows. Hence as ρ rises and MM is shifted rightwards, the gap narrows until the anticipated rate of inflation ρ equals the actual rate. In Figure 1 equilibrium occurs at the full-employment income level F where $MM(\rho>0)$ intersects EE at b and ρ equals $M'G'$; the real interest rate level Fb is less than the nominal rate which equals Fb plus ρ. It should be emphasized that if the elasticity of demand for money with respect to the nominal interest rate were lower, the MM schedule would have a steeper slope and the rightward shift of MM commensurate with the equating of ρ and the actual rate of inflation would be smaller. The equilibrium income level would then lie between L and F at a less-than full-employment level.[18]

It is now apparent that if monetary and fiscal policy had shifted the EE and MM schedules so that they intersected at a, the onset of induced expectations of inflation would lead to further changes in r and py. These considerations show that the previously discussed difficulties, befalling attempts to satisfy one or more of the three welfare objectives in the presence of exogenous inflationary or deflationary expectations, are compounded a fortiori with the onset of induced inflationary or deflationary expectations.[19] The only exception is that by engineering an induced deflation at the appropriate rate it may now be more feasible for the monetary-fiscal authorities to achieve full-liquidity, or at least to approach it more closely. Also, unlike the earlier hypothesized, exogenous onset of anticipated inflation, these considerations put the horse before the cart in so far as it is more realistic to presume that actual, or realized, rates of inflation induce the expectation of inflation. Again, this induced anticipated rate of inflation has the welfare benefit of giving the system a higher equilibrium level of output and employment—possibly even a full-employment equilibrium.

III. *The Payment of Interest on Money and the Welfare Objectives*

The above results indicate that if the monetary-fiscal authorities want to pursue the full-liquidity objective by inducing

[17] This is exactly analogous to the condition that ex ante investment must equal ex ante saving in equilibrium.

[18] At such a less-than full-employment equilibrium it should be emphasized that the nominal money wage level is changing at the same percentage rate (equal ρ) as the price level. This means that the ratio of the two levels, or the real wage, is constant at a level higher than that necessary for the achievement of a full-employment equilibrium. In this sense, the inflexibility or rigidity of the money wage refers to the fact that at any point in time it may be too high relative to the price level to give a full-employment equilibrium; hence under-employment equilibrium may be possible.

[19] When discussing this type of static short-run equilibrium analysis there is a strong temptation to "talk" as if one were doing dynamic analysis. For instance, this type of casual dynamics would describe the rightward shift in the MM schedule of Figure 1, consequent to the onset of the expectation of inflation, as if it were taking place over time. This in turn might lead one to conclude *erroneously* that the monetary expansion in excess of real product expansion, given by the distance MG, will itself shift the curve MM to the right. This is wrong. The static analysis represented in Figure 1 says nothing about the time path followed when the system moves from the disequilibrium position a to the equilibrium position b. All that it says is that if the percentage *rate* of change (equal $M'G'$) of the price *level* p associated with the nominal income *level* F is to equal the anticipated percentage *rate* of change (equal ρ) of the price *level*, then the equilibrium position of the system is at b; any other position, such as a, in not an equilibrium position because the actual rate of change of the price level is not equal to the anticipated rate. For a further discussion of this type of issue see Mundell's penetrating article.

TABLE 2—THE EFFECTS ON REAL OUTPUT, y, AND THE REAL INTEREST RATE, r, OF OPEN MARKET OPERATIONS AND CHANGES IN THE NET TAX LEVEL, T, ASSUMING INTEREST IS PAID ON MONEY

	dT	$d\lambda$	dM
dr	$\frac{dr}{dT} < 0$	$\frac{dr}{d\lambda} > 0$	$\frac{dr}{dM} \gtrless 0$*
dy	$\frac{dy}{dT} = 0$	$\frac{dy}{d\lambda} = 0$	$\frac{dy}{dM} > 0$

* $$\frac{dr}{dM} \gtrless 0$$

when

$$-S_w\left[\frac{\lambda c}{r}+k\right]-S_y \gtrless 0$$

as shown in the Appendix.

deflationary expectations, they would very likely have to be willing to incur a higher unemployment rate concomitant with a generally lower level of output and resource utilization. Pursuing the full-liquidity objective in this fashion would seem undesirable from a welfare standpoint. While the opportunity cost of holding money is being eliminated, or reduced, the aggregate output to be shared by all is smaller, i.e., the welfare of some is increased while that of others must necessarily be reduced. What of the feasibility of achieving the welfare goals by use of monetary-fiscal policy in conjunction with the payment of interest on money at a rate just sufficient to make the opportunity cost of holding it nil?

Assume that the appropriate interest rate is always paid, therefore that the full liquidity objective is always achieved; then the relevant demand equation for money is (4′), i.e., $m=ky$. Using this specification in place of (4), the whole system is totally differentiated with respect to the parameters under the control of the monetary-fiscal authorities; the relevant results (derived in the Appendix) are presented in Table 2.

Table 2 reveals that the possibility of simultaneously achieving both the interest rate and full-employment objectives is greatly enhanced in the case where sufficient interest is paid on money to achieve the full-liquidity objective. For now a change in the budgetary surplus T causes a change in the real interest rate r, in an unambiguous direction, but causes no change in real output y.[20] On the other hand, open market operations still affect both y and r, the former in an unambiguous manner, the latter possibly so. These considerations suggest that open market operations could be used to move real output y toward a full-employment position while changes in the budgetary surplus T could be used to adjust r toward the optimum real interest rate level.

If the government pays interest on money (with money), this will lead to a rate of change of the money supply in the same manner as does a non-zero value of T. Hence the rate of change of the money supply will be the joint result of both of these factors. As a consequence of this, the GM schedule of Figure 2 will no longer be parallel to the EE schedule as was the case in Figure 1. There GM was parallel to EE because the rate of growth in the money supply was solely the result of a budget deficit and the fact that the tax was lump sum. To this source of growth in the money supply now add that stemming from the payment of interest on money at a rate sufficient to make the opportunity cost of holding it nil.

Suppose at the real interest rate r'' (Figure 2) that the rate of deflation $G''M''$ is just equal to r'', and is therefore just sufficient to make the opportunity cost of holding money nil. Hence the government

[20] This seemingly counter-intuitive result, as well as the fact that $dy/d\lambda$ equals zero, will be explained shortly.

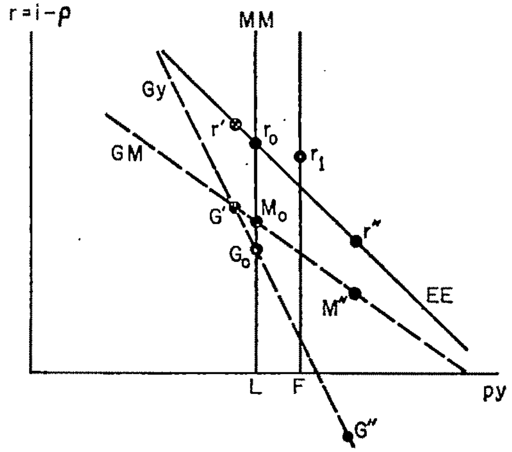


FIGURE 2

need not pay interest and the only cause of the rate of change in the money stock is the budget deficit T. Given T, if the government never paid interest on money, no matter what the nominal rate, the GM schedule would be parallel to the EE schedule, lying a distance $M''r''$ below it at all real interest rate levels, as was the case in Figure 1. At successively higher values of r the rate of deflation is successively smaller than the real rate of interest even when the only source of growth in the money supply is T. Hence if the opportunity cost of holding money is to be nil at real rates successively higher than the real rate r'', the government will now have to pay successively higher rates of interest on money, and this will, for given T, lead to successively higher rates of growth of the money supply. Because of this, at real interest rate levels above r'', the vertical distance between EE and GM will get ever larger.

At real interest rate r', the Gy and GM schedules intersect at G' and hence the rate of deflation will be zero; above r' the rate of growth of the money supply will exceed that of real output and there will be inflation. Whatever the *given* real rate above r', if the government attempts to pay interest on money in order to keep the opportunity cost of holding it at zero, this will only lead to a more rapid rate of growth of money, hence a higher rate of inflation, and, for the *given* real rate of interest, a higher nominal rate of interest which will necessitate still another increase in the government payments of interest on money, and so-on in an explosive spiral. Therefore, to the left of G' the GM schedule is not really defined because the system becomes explosive. These considerations suggest that even when the government pays interest on money in order to make the opportunity cost of holding it nil, it is still necessary to have continual deflation. This is because the system is only stable at real interest rate levels below r'.

Because the demand for money is now given by $M = kpy$ and there is only one value of py which will equate demand and supply in the money market, *given* the nominal supply of money, the MM schedule will be perfectly vertical. Suppose the nominal supply of money is such that this value of py is L in Figure 2, to the left of the full-employment income level F. The equilibrium real rate of interest is given by the intersection of the MM and EE schedules at r_0 and the nominal rate will be equal to the real rate Lr_0 minus the rate of deflation G_0M_0.

Suppose that, if the full-employment income level were reached, the optimum real rate of interest would be r_1. Table 2 indicates that an open market purchase would shift the MM schedule rightwards toward F, while at the same time shifting the EE schedule rightwards or leftwards, depending on whether dr/dM is greater or less than zero. It would be very unlikely that this operation alone would give rise to the optimum real rate of interest r_1 along with the achievement of full-employment. Consequently, in order to achieve r_1, given the achievement of F through the open market operation, the necessary ad-

justment could be made by changing the net tax level T without disturbing the previously attained full-employment money income level F, as can be seen from Table 2.[21] The resulting equilibrium position, characterized by the simultaneous achievement of r_1 and F, is only possible if the associated intersection of Gy and GM is to the left of the vertical MM schedule cutting through r_1 and F, that is, only if there is no inflation. Otherwise the system will explode as argued above.

This possible impediment to the simultaneous achievement of all three goals stems from the fact that the government is assumed to pay interest on the money supply. In a *laissez-faire* banking system in which it is legal to pay interest on deposits, profit maximizing banks would do so, but this would not lead to an increase in the money supply. The smaller the portion of the money supply on which the government would have to pay interest, the smaller would be the rate of growth of the money supply stemming from this source, and the more nearly parallel would the GM schedule be to the EE schedule. In the limit it would not matter if the rate of growth of the money supply exceeded that of output. For if the payment of interest on money were done entirely by private banks there would be no increase in the money supply from this source, hence the explosive situation discussed above would not develop. As this limit is approached, i.e., the larger the portion of the money supply on which interest payments are made by the private sector, the further to the left of the MM schedule in Figure 2 will the intersection of the Gy and the GM schedules occur—therefore the greater the likelihood that all three welfare objectives can be achieved simultaneously.

[21] It should now be clear why dy/dT and $dy/d\lambda$ both equal zero in Table 2. It is because the MM schedule is perfectly vertical and changes in T or λ only affect the EE schedule.

IV. *Summary and Conclusion*

When less-than full-employment equilibrium positions are possible, anticipated inflation is not unambiguously bad. Indeed the anticipation of inflation itself will be sufficient to move the system closer to a full-employment equilibrium position.

At the same time however, the shortfall from full-liquidity will be increased. Conversely, inducing deflationary expectations in an attempt to reduce the shortfall from full-liquidity will tend to increase the unemployment rate.

Moreover, we have seen that there is little reason to believe that the traditional tools of monetary-fiscal authority can achieve more than one of the three welfare objectives—full-employment, full-liquidity, or the optimum real rate of interest. Taking account of the *rate* of growth of output and the money supply, which are also affected by monetary-fiscal policies, and which serve to induce inflationary or deflationary expectations, makes the prospects even less sanguine.

Curiously enough, by making the opportunity cost of holding money nil through the payment of interest on money, the feasibility of achieving all three welfare objectives simultaneously is enhanced. Hence, in addition to the more familiar full-liquidity arguments for paying interest on money, the enhanced feasibility of simultaneously achieving full-employment and an optimum real rate of interest made possible by such payments lends further justification to their implementation.

APPENDIX

The model discussed in this paper may be reduced to three simultaneous equations by substituting equations (2) and (5) into (3); the resultant expression for w is then substituted into equation (6). The variable N is eliminated from the system by reformulating the production function $y=y(N)$ as

$N=y^{-1}(y)$, and hence the labor market equilibrium condition

$$(1') \quad \frac{\overline{W}}{p} = y_N(N) = y_N(y^{-1}(y)) = g(y)$$

Now $y^{-1}(y)>0$, $y_N(N)>0$, and $y_{NN}(N)<0$. Therefore $g(y)>0$ and $g_y(y)<0$. Equation (1) is then replaced by (1′). The resulting three equation system is

$$I(r) - S\left(r, \lambda\frac{cy}{r} + \frac{\overline{M}}{p}, y, \overline{T}\right) - \overline{T} = 0$$

$$\frac{\overline{M}}{p} - yk(r) = 0$$

$$\frac{\overline{W}}{p} - g(y) = 0$$

By totally differentiating this system and solving the resulting equations for

$$\frac{dr}{dT}, \frac{dr}{d\lambda}, \frac{dr}{dM}, \frac{dy}{dT}, \frac{dy}{d\lambda}, \text{ and } \frac{dy}{dM}$$

we have

$$\frac{dr}{dT} = \frac{1}{\Delta}(S_T+1)(k(r)\overline{W} - g_y\overline{M})\frac{1}{p^2} < 0$$

$$\frac{dr}{d\lambda} = \frac{1}{\Delta}(k(r)\overline{W} - g_y\overline{M})\frac{a}{p^2}S_w > 0$$

$$\frac{dr}{dM} = \frac{1}{\Delta}\left[S_y + S_w\left(\lambda\frac{c}{r} + k(r)\right)\right]\frac{\overline{W}}{p^3} \gtrless 0$$

$$\frac{dy}{dT} = -\frac{1}{\Delta}(S_T+1)\frac{\overline{W}}{p^2}yk_r < 0$$

$$\frac{dy}{d\lambda} = -\frac{1}{\Delta}aS_wyk_r\frac{\overline{W}}{p^2} > 0$$

$$\frac{dy}{dM} = \frac{1}{\Delta}\left[(I_r - S_r) + S_w y\left(\frac{\lambda c}{r^2} - k_r\right)\right]\frac{\overline{W}}{p^3} > 0$$

assuming $-1<S_T<0$, $g_y<0$, $S_w<0$, $S_y>0$, $k_r<0$, $I_r<0$, and $S_r<0$. Δ is the basic determinant of the system and equals

$$-\frac{S_w}{p^2}\left\{\lambda\overline{M}g_y\frac{cy}{r^2} - \lambda\overline{W}\left[yk_r\frac{c}{r} + k(r)\frac{cy}{r^2}\right] - yk_rg_y\overline{M}\right\} - \frac{\overline{M}}{p^2}(I_r - S_r)g_y + \frac{\overline{W}}{p^3}[(I_r - S_r)k(r) + S_yyk_r]$$

$\Delta<0$ if we assume

$$-\overline{M}g_y + k(r)\overline{W} > -r\overline{W}k_r$$

which seems reasonable given that $g_y<0$ and given that the nominal money stock $\overline{M}$ is typically many times the size of the nominal wage $\overline{W}$; $dr/dM \gtrless 0$ when

$$-S_w\left[\frac{\lambda c}{r} + k(r)\right] - S_y \gtrless 0$$

When interest is paid on money balances at a rate sufficient to make the opportunity cost of holding it nil, equation (4′) is the relevant demand equation for money. The results of totally differentiating the whole system are the same as above except that k_r now equals zero. Hence dy/dT and $dy/d\lambda$ now equal zero. Also, the basic determinant of the system is unambiguously negative.

REFERENCES

M. J. Bailey, "The Welfare Cost of Inflationary Finance," *J. Polit. Econ.*, Apr. 1956, *64*, 93–110.

M. Friedman, "Discussion of the Inflationary Gap," reprinted in *Essays in Positive Economics*, Chicago 1953, pp. 253–57.

———, *The Optimum Quantity of Money and Other Essays*, Chicago 1969.

A. Leijonhufvud, *On Keynesian Economics and the Economics of Keynes*, Oxford 1968.

L. G. Metzler, "Wealth, Saving, and the Rate of Interest," *J. Polit. Econ.*, Apr. 1951, *59*, 93–116.

R. A. Mundell, "A Fallacy in the Interpretation of Macroeconomic Equilibrium," *J. Polit. Econ.*, Feb. 1965, *73*, 61–67.

E. S. Phelps, "Anticipated Inflation and Economic Welfare," *J. Polit. Econ.*, Feb. 1965, *73*, 1–17.

Constraints on Public Action and Rules for Social Decision

By David F. Bradford*

Several years ago, in an unpublished note, Robert Dorfman presented two formulas for averaging individual rates of time preferences to obtain a discount rate to use in deciding whether to reject or accept a public investment opportunity. Essentially both formulas are based on the same compensation test, with the difference that one uses individual present values to express gains and losses, while the other uses the equivalent annual perpetuity streams. There is apparently complete symmetry between the two formulas; neither approach is obviously more convincing for public investment choice than the other. Yet the two formulas may, and usually will, yield different values for what may be labeled the "project-specific social discount rate,"[1] and will therefore lead to different social investment choices. I refer to this as the "Dorfman paradox."

This paper presents an analysis of the Dorfman paradox, using it as a vehicle for exploring the prescriptive force of benefit-cost techniques. Benefit-cost analysis attempts to discover what quantities of some specified numeraire good (usually current dollars) are considered by the gainers and losers from some proposed government action to be just equivalent in value to their respective gains and losses. The algebraic sum of these quantities of the numeraire good over all individuals is then used to assist in deciding whether to undertake the action. We say that a project "passes the benefit-cost test" if this net sum is positive. There is at least a presumption that projects passing the benefit-cost test should be undertaken, and those failing should be rejected.

The apparatus of welfare economic theory which lies behind this procedure builds upon a foundation of competitive equilibrium in the production and exchange of private goods. Under this assumption it makes no difference whether "apples" or "pears" are chosen as numeraire for analysis of the public action, for an analysis in terms of pears can be translated into one in terms of apples simply by multiplying all pear quantities by the equilibrium apple price of pears.

When, for some reason, competitive equilibrium in the markets for private goods has not been attained, so that individuals differ in the rate at which they are willing to exchange apples for pears, the choice of numeraire good may have a significant bearing on the results of a benefit-cost analysis. In particular, a change in the choice of numeraire may lead to a change in the net aggregated value of an action from positive to negative, or vice versa. This is the essence of the Dorfman paradox, which arises in the case of great practical significance, where

* The author, assistant professor of economics at Princeton University, wishes to acknowledge contributions by Charles L. Tucker and Thomas Latham, former graduate students at the Woodrow Wilson School of Public and International Affairs at Princeton. Conversations with them on the topic helped considerably to sharpen the analysis. He would also like to thank Robert Dorfman and a referee for comments on an earlier draft, including suggestions which have enriched and extended the analysis. Needless to say, the author bears sole responsibility for opinions and errors.

[1] It is called project-specific because the rate calculated to apply to a particular investment opportunity depends upon the identity of the beneficiaries or of the losers associated with that project.

a competitive equilibrium does not hold in the production and exchange of dated goods. In more familiar terms this is the case of imperfect capital markets.

While the immediate occasion for this paper is the unraveling of the Dorfman paradox, my intent goes beyond this to two broader objectives. First, it is shown, in a carefully circumscribed model, what is the "force" of benefit-cost analysis applied to choice of social actions. Why *should* a project passing a benefit-cost test be undertaken? In answering, I stress the importance of being aware of any existing constraints on social action in designing and using a benefit-cost test.

The second objective is to contribute to an understanding of the appropriate use of non-equilibrium prices in benefit-cost analysis. This is a relatively undeveloped area of theory, although the proper (shadow-) pricing of unemployed resources has been frequently discussed, and some thought has been given to the problem of valuing resources bid away from, or output purchased from, a monopolist. Here again, I conclude that an understanding of the actual constraints on social action is essential to the design and use of benefit-cost techniques.

Although the use of non-equilibrium prices, especially non-equilibrium interest rates, in benefit-cost analysis merits study on its own, the case for undertaking this analysis of the Dorfman paradox is strengthened by the existence of at least one important study in which one of the Dorfman tests was implemented. In *Multiple Purpose River Development*, John V. Krutilla and Otto Eckstein use an average of individual discount rates, weighted by tax shares, in calculating the net present value of a public investment project. The analysis contained in this paper shows what must be assumed for their procedure to be an appropriate basis for social choice.

The extensive literature on the social discount rate makes clear that conclusions are very sensitive to the choice of models on which they are based.[3] In the next section, I describe a formal model of benefit-cost analysis, and in Section II, discuss the case of social investment choice with perfect capital markets. In Section III the Dorfman paradox is presented and analyzed in terms of the implicit constraints on social action. Section IV contains a summary and concluding remarks. In the Appendix, I discuss the effect on the analysis of the use of weights to reflect distributional objectives.

I. *The Model World*

The analysis is confined to a world of certainty in which a benevolent but not omnipotent dictator attempts to maximize a Bergsonian-Samuelsonian welfare function. Its arguments are utility indicators of the members of the (static) population. The arguments of the individual's utility function, in turn, are his consumption of the economy's single private good in the present period, his future consumption rate (restricted for simplicity to the form of a constant perpetuity), and the future level of output of some government-provided public good or goods. The dictator influences welfare by his choice of "social actions," where a social action is described by a detailed list of its effect on the arguments of each individual utility function. Examples of social actions are pure transfers and investment projects. In a typical transfer, Peter's present consumption is reduced by one unit and Paul's is increased by one unit. In a typical investment project Peter's present consumption is reduced by $\frac{1}{4}$ unit and Paul's by $\frac{1}{4}$ (to build a bridge), and the item, "future services of one bridge," is entered in the accounts of all citizens.

[3] Cf the Marglin-Tullock-Lind and the Eckstein-Hirshleifer exchanges.

By his private investment decision an individual can alter his combination of claims to present and future consumption of the private good according to a smooth technological transformation function which exhibits the usual diminishing marginal rate of return on foregone present consumption. Given the values of present and future consumption levels and the level of output of the public good, the minimum perpetual increment to future consumption which could be substituted for one dollar of present consumption to leave the individual no worse off is called his marginal rate of time preference. I assume that none of the changes discussed below are sufficient in magnitude to alter this trade-off rate and therefore label it simply his time preference rate. If all possibilities for profitable production and exchange of claims to the private good are exploited, all individuals will have the same time preference rate, which will equal the marginal rate of return on all transformation functions.

Just as we can calculate an individual's time preference rate if we know the values of the arguments of his utility function, it is assumed that we can also calculate the maximum amounts of present or future dollars of the private good which he is willing to give up to obtain specified levels of provision of public goods. Of course, if r_i is the ith individual's time preference rate, and x_i is the maximum future perpetuity payment he would be willing to make to obtain some public service, then the maximum present payment he would make for the same service is x_i/r_i. I call the first quantity, x_i, the *perpetuity value* of the service to him, and the second quantity, x_i/r_i, its *present value.*[3]

[3] As was pointed out by the referee, the word "value" in such discussions may be misleading. The Dorfman paradox arises precisely when we do *not* have a market *value* to deal with, but only the set of differing individual marginal rates of substitution.

II. *Benefit-Cost Tests with Perfect Capital Markets*

We may now ask, in this model, what use the dictator might make of benefit-cost studies of the available public investment projects under the assumptions of competitive equilibrium in the private goods market. As mentioned above, the term "investment project" connotes somewhat more than it does usually. A public investment project is described as usual by the total amount of the private good which must be given up in the present, and the yield in the form of a measure of the perpetuity level of provision of the public good in the future. The project description includes, in addition, the identities of those individuals from whom the present resources will be taken to finance the project. Let us assume that the hard part of benefit-cost analysis, the estimation of the monetary value of benefits, is easily accomplished, for example, by direct interview of an honest and imaginative citizenry. We may then replace the measure of public good output in the project description with a specification of the private good perpetuity valuation placed on the public service by each member of society. Following Dorfman, let K be the total amount of the present private good which must be given up to finance a typical public investment, and let t_i be the fractional share of this sacrifice made by citizen i. Let b_iY be the perpetuity value of the public service to citizen i, where $\Sigma b_i=1$, so that the sum of these valuations over all citizens is Y.[4]

Usually, a benefit-cost test is constructed by discounting the aggregate benefit stream, Y, at some appropriate

[4] Throughout this paper I assume that investment projects yield present-period gains to no one and impose future-period losses on no one. This assumption has no material effect on the argument; negative values of t_i and b_i could be incorporated to it directly. I have followed Dorfman in posing the benefit-cost

rate and comparing the result with K. Under the assumption that there are no externalities to (marginal) private investment, and that only the utility of existing citizens counts, the "appropriate rate" by which future gains should be discounted, according to the usual arguments of welfare economics, is the single market rate of interest, r. If Y/K exceeds r, the project "passes" the test.

It is of some interest to identify the circumstances under which the dictator can delegate some of his work to the staff level through the use of benefit-cost analysis. The dictator's decision-making load might be reduced by screening out projects he would definitely reject. Since he is assumed to be solely equipped to evaluate social welfare, the procedure should not remove from his field of choice of social action any which is not dominated by another action available to him. (An action A is said to dominate action B if all citizens weakly prefer A to B and at least one strongly prefers A to B; i.e., A is Pareto-superior to B.) Still better, the analysis might equip the staff to make positive decisions, incorporating sufficient as well as necessary conditions for acceptance. The dictator could in this case concern himself with transfer programs alone.

The power of the simple benefit-cost test described above to perform these functions may be expressed in the following two propositions:

Proposition 1. If the dictator has the power to make (costlessly) any direct transfer of the private good from one individual to another, any project which fails the benefit-cost test can be dominated by a feasible program of transfers. If a project passes the test, a program of transfers plus the undertaking of the project can be found which will dominate any program of transfers in the absence of the project. Hence, he can safely instruct his staff to undertake only such projects for which Y/K exceeds r, and need concern himself only with adjusting the welfare distribution through transfer payments.

The reasoning behind this statement, i.e., the manner of constructing the requisite transfer programs, will be familiar to every economist as that by which mutually advantageous exchanges are derived when different relative prices prevail in different locations between a pair of commodities, and should require no elaboration.[5]

Proposition 2. If the dictator can redistribute welfare only via public investment projects, then there may be no action available which dominates a project which fails the benefit-cost test, while a project which passes the test may not dominate the available alternative actions. Thus, the dictator may wish to undertake projects failing the test, and may reject projects passing the test, so he cannot delegate project selection through this device.

The argument behind this proposition is very simple: If there is no other feasible

tests in the form of ratios. If disinvestment projects are to be considered this could cause confusion, since $Y-rK\geqq 0$ implies $Y/K\geqq r$ only if K is greater than zero. The possibility of confusion here could be avoided by dealing always with inequalities of the former type, at the cost of some of the intuitive appeal of the presentation.

[5] The assumption that no change contemplated will be large enough to alter individual discount rates is critical to the strict validity of Proposition 1, and this in turn requires strictly that indifference curves be "flat," at least in the neighborhood of the point from which changes are contemplated. If the project in question can be constructed to any scale, then Proposition 1 will always be true for a "sufficiently small" amount of it, without any assumption about constancy of discount rates.

On the impossibility in general of using current prices to assess whether the beneficiaries from any given finite change in social totals of commodities can compensate the losers and still benefit from the change, see Samuelson.

action by which the dictator can achieve a gain for any given citizen, equivalent to that he would receive from a project, then the project cannot be ruled out of consideration a priori, for this means precisely that the action cannot be dominated. Thus, if the government's objective is to improve the lot of some specified group, such as the poor, and direct transfers are politically infeasible, apparently "unprofitable" investment projects may be desirable. Conversely, a nominally "profitable" project may benefit the rich at the expense of the poor. If the dictator cannot alter this outcome by appropriate transfers, he may consider that the adoption of the project will reduce social welfare. The matter can be settled only by reference to a social welfare function.

III. *The Dorfman Paradox*

In this section we shift from one extreme assumption—competitive equilibrium in the private capital market—to the other extreme—the complete absence of private exchange between different individuals of present for claims on future consumption. Assume that an individual can vary his pattern of present and future consumption only by varying his own real investment program. Each individual chooses an optimum combination subject to this constraint; denote by r_i his resulting rate of return on investment and marginal rate of time preference.[6]

As Dorfman pointed out, we may now quite naturally define two different benefit-cost tests for a social investment. The net perpetuity value of the project to citizen i is $b_iY-r_it_iK$; the net present value is b_iY/r_i-t_iK. The first test, then, the *present value test*, is that the sum of individual net present values of the project be greater than or equal to zero:

$$(1) \qquad \sum (b_iY/r_i - t_iK) \geq 0$$

Using the fact that $\Sigma t_i = 1$, this is equivalent to the condition:

$$(2) \qquad \frac{Y}{K} \geq \frac{1}{\sum \frac{b_i}{r_i}},$$

that is, that the rate of return on the project exceed the harmonic mean of individual time preference rates, weighted by benefit shares.

The second test, the *perpetuity value test*, requires that the sum of individual perpetuity values of the project exceed or equal zero:

$$(3) \qquad \sum (b_iY - r_it_iK) \geq 0$$

Here we use the fact that $\Sigma b_i = 1$ to derive the equivalent condition:

$$(4) \qquad \frac{Y}{K} \geq \sum t_ir_i,$$

i.e., the rate of return on the project must exceed the arithmetic mean of individual time preference rates, weighted by tax shares. For any given project, these two criteria may, and generally will, differ. The paradox lies in the fact that a project may pass one test and fail the other. From one point of view the project provides sufficient benefits for the gainers to profitably compensate the losers, while from another this same condition is not fulfilled.

One difference between the two tests is immediately apparent. Since a given present value gain or loss is equivalent to a larger perpetuity gain or loss the larger is the individual's rate of time preference, the present value test will tend to produce investment choices favored by the low time

[6] In fact we are here interested principally in the time preference aspect of the r_i's; the entire argument of the paper holds without any private investment possibilities. However, the assumption that the r_i's are unaffected by the government actions considered is more plausible if private investment is allowed. For example, if an individual can invest as much as he wants in a given process with a yield of 3 percent, we know his marginal rate of time preference will remain 3 percent under any "exogenous" changes in his claims to present or future consumption, as long as he invests anything at all in the process in question.

preference individuals, relative to the choices indicated by the perpetuity value test.[7] For suppose x_i to be the perpetuity value (positive or negative) of the project to individual i, as employed in the perpetuity value test. In the summation used in the present value test this quantity is "blown up" by the factor $1/r_i$, so that a given perpetuity gain or loss acquires the more weight in the present value calculation the lower is the individual's discount factor. Notice that, since x_i may be negative for individuals with a low discount rate, this does not imply that the critical rate by the present value test is always lower than that by the perpetuity value test for a given project. (I refer to the cut-off rate of return as reckoned by a given test as the "critical rate" for that test.) That is, the low discount rate people may be against the project, so that as they acquire more influence in the decision the cut-off rate tends to rise. (For an example see Table 1, which is explained below.)

A second immediately apparent difference between the tests is that in the present value test, tax shares and the time preference of taxpayers, qua taxpayers are ignored, while in the perpetuity value test benefit shares and the time preferences of beneficiaries are left out of the reckoning. A closer analysis reveals that the present value test is the correct instrument for project selection when transfers of present, but not future, dollars are unrestricted, and hence when tax shares are in fact irrelevant because they are not genuinely fixed. Similarly, the perpetuity value test is the appropriate decision criterion when future perpetuity dollars can be freely redistributed, but present dollars not.

These generalizations can be clearly demonstrated in a simple model world of only two individuals (or two classes of individuals), individual (or class) H having a "high" rate of time preference (and marginal rate of return on investment), r_H, and individual (or class) L having a "low" rate, r_L. Let b_H and t_H be the benefit and tax shares, respectively, of the first individual, b_L and t_L, those of the second. To investigate the choices indicated by the two tests, consider the consequences of using them under each of the four possible extreme distributions of taxes and benefits. Table 1 shows for each such distribution the computed critical rate of return which a project must reach in order to pass each test.

TABLE 1

Case	b_L	t_L	b_H	t_H	Present Value Test Critical Rate	Perpetuity Value Test Critical Rate
1	1	1	0	0	r_L	r_L
2	1	0	0	1	r_L	r_H
3	0	1	1	0	r_H	r_L
4	0	0	1	1	r_H	r_H

The entries in Table 1 are calculated according to the formulas derived above. For example, in tax-benefit distribution Case 2, the low discount rate individual receives 100 percent of the benefits, which occur in the future, while the high discount rate individual pays all of the taxes to make the investment in the present. The present value test is thus

$$\frac{Y}{K} \geqq \frac{1}{\frac{1}{r_L} + \frac{0}{r_H}} = r_L$$

The perpetuity value test is

$$\frac{Y}{K} \geqq 0 \cdot r_L + 1 \cdot r_H = r_H$$

Let us suppose that a project being contemplated yields a rate of return, Y/K or r_m, which is intermediate between r_L and r_H. The two tests are in agreement; in Case 1, that the project passes, and in Case 4, that it fails. This is easily inter-

[7] This was first pointed out to me by Charles L. Tucker and Thomas Latham.

preted, since in the former case there is no loser and in the latter case, no gainer, if the project is undertaken. In Cases 2 and 3, the tests give contradictory results. The interpretation is again not difficult however. In Case 2, the project passes the present value test because the gain to individual L due to the project exceeds that obtainable by a simple transfer of K present dollars from individual H. The project fails the perpetuity value test under the Case 2 gain-loss distribution because it can be dominated by a perpetuity transfer of any amount between r_mK and r_HK from H to L (since any such transfer is at least as good as the perpetuity benefit Y to L yielded by the project, and no worse than the present sacrifice of K by H).

Similarly, under gain-loss distribution 3, the project fails the present value test, since it is dominated by a present transfer of any amount between Y/r_H and K from L to H, and it passes the perpetuity value test, since there is no perpetual future transfer from L to H which dominates it.

The results of this analysis of Dorfman's paradox can be summarized in a series of propositions, analogous to the two of the previous section, to apply to the model world with imperfect capital market:

> *Proposition* 1′a. If the dictator can make any transfer he desires, either of present dollars or of perpetuities, from one person to another, a transfer program will exist which dominates any investment project with a rate of return less than the *maximum available rate* (in the case above, r_H), while if a project has a rate of return in excess of that, there will be a feasible program of transfers plus the undertaking of the project which dominates any pure transfer program.

In effect, the ability to make any transfer should enable the dictator to settle for nothing less than Pareto optimality in the distribution of private goods. Unequal discount rates across individuals signal, under our assumptions, the opportunity for gains from trade. Even if our dictator does not push these gains to the limit, there will always be a transfer of present dollars to, and perpetuity future dollars away from, individual H which dominates any project with return less than r_H.

> *Proposition* 1′b. If the dictator can transfer present dollars from one citizen to another, but not future perpetuities, any project which fails the present value test
>
> $$\left(\frac{Y}{K} \geq \frac{1}{\sum b_i/r_i}\right)$$
>
> can be dominated by a feasible program of transfers, and there will be a feasible program of transfers plus a project passing this test which will dominate any pure transfer program.

> *Proposition* 1′c. If the dictator can transfer future perpetuities from citizen to citizen, but not present dollars, there will be a feasible transfer program which dominates any project which fails the perpetuity value test ($Y/K \geq \Sigma t_i r_i$), and there will be a feasible program of transfers plus a project passing this test which dominates any pure transfer program.

Propositions 1′b and 1′c simply state that if a numeraire good is the only good transferable by the dictator, a project whose net value in terms of the numeraire is positive cannot be bettered by a simple transfer of the numeraire, while a project whose net value is negative can be. Propositions 1′a–1′c "collapse" to Proposition 1 if there is competitive equilibrium in the market for perpetuities.

> *Proposition.* 2′. If the dictator can redistribute welfare only via public invest-

ment projects then no test (other than that there be no gainer) will indicate that a particular project can be dominated by feasible alternative action, and no test (other than that there be no loser) will indicate that a particular project will always be included in a program of actions by the dictator in his attempt to maximize welfare. That is, no tests either for rejection or acceptance are available for decentralized decision-making.

Dorfman's paradox arises because the numeraire used to measure value does not have the same equivalent in other commodities for all individuals. The paradox is here developed in an essentially two-commodity world; in a multiple commodity world (e.g., one with diverse term structures of interest rates) it would lead to multiple "tests," one corresponding to each potential numeraire good (which could be dollars in different future periods or different present commodities) for which more than one exchange rate with other commodities is observable. In each case the test would indicate whether a project should be undertaken under the assumption that the designated numeraire good, and no other goods, can be redistributed.

IV. *Summary and Concluding Remarks*

The limitations of hypothetical compensation tests have been thoroughly examined in the literature. That benefit-cost analysis as a guide to social action amounts to the use of a hypothetical compensation test has also been recognized. Such analysis is generally accompanied by the caveat that the goodness or badness of an action's effect on the income distribution must be considered along with net benefit or benefit-cost ratio calculations in reaching a social decision. This has two consequences. On the one hand, an action yielding positive net benefit may sometimes be rejected if it has an unfavorable impact on the income distribution, so that benefit-cost analysis will not provide a test for acceptance. Conversely, it may sometimes be desirable to accept actions with negative net benefits if the impact on the income distribution is salutary, and hence benefit-cost analysis will not provide a test for rejection. That is, limitations on the government's ability to make direct transfer redistribution of income prevent benefit-cost analysis (even assuming it is feasible to express *all* gains and losses in terms of a common numeraire) from being a sufficient basis either for definitely accepting or definitely rejecting any specfic social action. These observations are formalized in Propositions 1 and 2 above.

When all individuals are in agreement about the rate of exchange between present and future consumption, as in the case of perfect capital market equilibrium, it doesn't matter whether the summation of private net gains used in the analysis is carried out in present consumption or future consumption units. Either way amounts to using *the* private discount rate as a social discount rate.[8] However, it does matter which good is used as numeraire when there is a variety of private discount rates, presumably signalling capital market disequilibrium, which is why Dorfman's two methods for combining them to generate a social discount rate yield different answers. Which if either of these methods is appropriate turns out to hinge on the particular type of income transfers considered feasible. These con-

[8] It might have been better to have avoided the term "social discount rate," given the many unresolved theoretical controversies surrounding it. However, notice that to focus on the problem at hand, I have side-stepped many of these issues by the specification of the model, especially in assuming away uncertainty and future generations. In principle, future generations could be incorporated directly by assigning individuals who will first enter perpetual existence "tomorrow" very low, in fact, zero r_i's.

clusions are stated formally as my Propositions 1′a, 1′b, 1′c and 2′.

What the analysis makes clear is that when there is such a diversity of rates, if not all redistributions are possible, it is not proper to speak of *a* social discount rate for evaluating all social investment projects. If present, but not future, transfers are possible, the time preferences of the beneficiaries of a particular project (more generally, of those whose claims to future private or public good services are affected) are relevant for a decision, while the time preferences of the losers (more generally, of those whose claims to present services are affected) need not be considered. The pattern of beneficiaries varies from project to project, and along with it the discount rate appropriate for decision. If future, but not present, transfers are possible, time preferences of the beneficiaries become irrelevant, while the time preferences of the losers become relevant since these determine whether a project can be dominated by future transfers. The pattern of losers, and hence the project-specific discount rate, may also vary from project to project. Of course, to the extent that a single form of finance is used to provide resources for a wide class of social investment projects, a single rate may be calculated to apply to all in the class. This underlies the procedure followed by Krutilla and Eckstein, who calculate the perpetuity-value test discount rate, which they call the social cost of capital, for various sources of finance. The resulting rate is used to evaluate multiple purpose river development projects. Krutilla and Eckstein do not, however, justify their procedure on the basis of either limitation on present transfer possibilities or availability of future transfer possibilities.

Of the two combinations of direct transfer possibilities—present possible, future not possible, or present not possible, future possible—the latter is perhaps the one more likely to approximate reality, for at least two reasons. First, limitations on transfer possibilities are likely to be institutional, and institutions permitting present transfers would seem likely to permit future transfers. If present institutions do not permit some kinds of transfers, one can still hope or expect that future institutions will. Second, a society gets "two shots" at future transfers. In the discussion above "future transfers" may be taken to mean present transfers of claims to future income; however, future transfers can also be brought about by actual transfers of (then present) income made in the future.

The dissection of the Dorfman paradox reveals that there need be no ambiguity about the "correct" way to incorporate into benefit-cost analysis a multiplicity of interest rates arising from imperfection in the capital market.[9] The choice of test is dictated by the constraints on social action, in this case on the power to carry out certain types of transfer. It must be recognized, of course, that the happy possibility of reducing social decision criteria to simple formulas is much less likely to occur where the constraints on social action are more selective or complex. We have seen how relatively unconstrained must be the government's transfer power in order for *either* of the Dorfman tests to fulfill a positive (sufficient condition) decision function, or even a screening (necessary condition) function, in project selection. At the other extreme, where the dictator is assumed unable to make *any* explicit

[9] In this analysis of the Dorfman paradox no attempt has been made to provide an explanation of the diversity of observed rates of time preference, simply attributing them to the absence of exchange. The same analysis would, of course, apply to a system of separated capital markets, each separately in equilibrium but with no exchange among sub-markets. Whether or not the precise source of interest rate diversity (monopoly, transactions costs, etc.) has implications for the analysis is a question awaiting further work. Note, however, that the rate differences are assumed *not* to reflect differences in risk or liquidity; an interest rate is here a rate of exchange or trade-off between a definite pair of goods, present consumption and future perpetuity consumption with certainty.

transfers, developing formal decision rules through the device of weights reflecting his tastes in income distribution is a possibility, although perhaps only a formal one.[10] Whether things are simpler in the intermediate range can only be determined by case to case investigation.

In drawing policy implications one must not only observe that real governments' transfer capabilities are limited, but also that a democratic government is probably not very well modelled by a benevolent dictator equipped with welfare function. The first fact suggests that for any given project the simple information that it passes a benefit-cost test does not insure that it will be worked into the program of a "good" government. Furthermore it suggests that benefit-cost tests are probably also not suitable screening devices, since a good government will want to consider projects which fail such tests. Relevant to the decision are not only the totals, the net gain or loss summed over the citizenry, but also the fine detail, identifying the gainers and losers. The inclusion of such information in benefit-cost analysis has been urged by a number of writers on the topic,[11] and there has been a trend in recent years in the United States, at least at the Federal level, toward putting this suggestion into practice, especially in the analysis of programs with predominantly redistributive intent. The present argument supports this procedure.

A democratic government is unlike our benevolent dictator not only in the fact that its decisions have little claim to a higher ethical sanction,[12] but also in the fact that its decisions are much less likely to display the same consistency as his.[13] If we restate the objective of the benefit-cost analysis to be the prediction of whether a project will be undertaken by the government, and sever its connection with the assessment of the project's ethical value, we get rid of the first problem, but not the second. It is at least debatable whether the taste of a legislature for income redistribution can be satisfactorily represented by weights of the type discussed in the Appendix of this paper.[14]

It should not be overlooked that one of the important functions of the politician is precisely to discover and organize those exchanges which cannot or are not brought about directly by those interested, as in the private market place. For this purpose, both the information that a benefit-cost test is passed by a project and the detailed identity of gainers and losers are of interest. The first tells the politician that there is at least the possibility of arranging matters so that the project is undertaken and everybody is better off (surely every politician's dream). The second, of course, helps him to bring about a successful exchange, or at least to look after the interests of his constituents. Needless to say, the politician will reach his conclusions about these matters even if the economist does not provide him with the detailed information.

Two closing remarks are germane to the analysis in this paper, but unconnected with the immediately preceding discussion. The first is to point out that while the constraints on the dictator's action explic-

[10] See the Appendix to this paper.

[11] Burton A. Weisbrod, who is to be numbered among these writers, documents several of these recommendations (esp. p. 179, fn. 2).

[12] A comment by the referee on another part of the paper makes this point rather succinctly; he writes: "Just pronouncing the incantation 'Bergson-Samuelson social welfare function' hardly guarantees the ethical quality of government decisions. The answer, I'm afraid, is that we must consider the solutions equally indeterminate ethically—whether we look at the market process or the political process of human interaction."

[13] It seems hardly necessary in this connection to cite the work of Kenneth J. Arrow which prompted extensive modern research into the nature of social welfare as well as into the choices generated by various political mechanisms.

[14] For an optimistic view on these matters, see Weisbrod, and for a forceful statement of the sceptic's position, see the comments on Weisbrod's paper by Ruth P. Mack in the same volume, esp. pp. 216–22.

itly considered in the simple model world of Dorfman's paradox apply to the making of transfers, a highly significant constraint implicitly assumed is that he is unable to carry out profitable arbitrage in the capital market, even though recognizing the existence of different prices. "If he's so smart, why ain't he rich?" In situations where the analyst is convinced he is dealing with different prices for the same good it is to be hoped that he will not forget to explore the reasons that equilibrating exchange does not take place and the potential for social profit in arbitrage operations by the government.

The second remark is that, although the limited ethical content of compensation tests has been pointed out, and appropriate scepticism about their use in government decision making has been expressed, the fact remains that in the market sector a compensation test, in the guise of the test of profitability, is determining a very significant share of allocative decisions of the economy. This prompts the question whether economists are being too hard on compensation tests in government decisions or too soft on market allocations. Is it reasonable to assume implicitly that the government will intervene (through transfers?) to maintain a "good" wealth distribution after allocative decisions have been made on pure benefit-cost test grounds in the private market, but not after the corresponding decisions within the government sector? It is possible that the answer to this question is yes, but, to my knowledge, the reasoning has yet to be spelled out by writers in this area.

Appendix

Weighting Costs and Benefits[15]

A device often suggested by which evaluation of distributional considerations might be introduced into project selection is the use of weights to reflect the social value of costs and benefits to different individuals. Such weights might be explicitly or implicitly specified by a legislature, for example. In our model world they would embody the dictator's judgment of the increments to the value of the social welfare function obtainable by giving a unit of numeraire good to different individuals.[16]

If the dictator is doing his job this device has a role to play only in the case in which his power to make transfers is limited. For if a dollar to Peter has a different social value than a dollar to Paul, and if he is not restricted from doing so, the dictator must carry out the transfer in the appropriate direction to increase welfare. He will stop making transfers when the marginal social value of a dollar to any citizen is the same for all; in other words, he will see to it that the weights are all unity.[17] However, when transfers are ruled out, or at least further transfers in directions which would increase the value of the welfare function, and the changes are reasonably small, weights would allow decentralized project selection.

In the world where the Dorfman paradox occurs, i.e., where the opportunities for mutually profitable exchange of dated goods are not exhausted, care must be excercised to be sure that the weights used are those appropriate to the numeraire in terms of which costs and benefits are measured, present dollars or perpetuity future dollars. For now to say that giving Peter \$1.20 of present income is equivalent, in its effect on social welfare, to giving Paul \$1.00 (i.e., Paul is more deserving), no longer implies

[15] The inclusion of a discussion of the use of weights in the cost and benefit calculations was suggested by Robert Dorfman in commenting on an earlier draft. He also first pointed out to me the symmetry of the formulas derived in this section.

[16] For a recent and sophisticated analysis of the use of such weights in project selection, see Weisbrod. In the cited work Weisbrod argues for the systematic inclusion of information about the identity of beneficiaries along with "efficiency" indicators in project selection, essentially because of limitations on the government's transfer capabilities.

[17] Strictly speaking, the process will be carried on until the *ratio* of any pair of weights is unity. The absolute value of the weights plays no part. However, it simplifies the presentation if we normalize the "weight level" at unity.

that a \$1.20 future annuity to Peter has the same social worth as a \$1.00 annuity to Paul.

The weights appropriate to the use of the two possible numeraire goods in our model world are related to each other in a simple way. Suppose that gains to citizen i measured in present dollars are given weight W_i, i.e., that W_j/W_i is the number of present dollars we must give to i to keep social welfare constant if we take one dollar away from j (above, $W_{\text{Paul}}/W_{\text{Peter}}=1.2$). If r_i is i's discount rate, the gift of a \$1 annuity to him has the same effect on welfare as a gift of $1/r_i$ present dollars. Hence the dollar size of the future annuity which we must give to i to keep social welfare constant if we take a \$1 annuity away from j is $(W_j/r_j)/(W_i/r_i)$.

Put another way, if a package of net gains g_i, in the form of present dollars, raises social welfare by $\Sigma\ W_i g_i$, then a package of net gains g_i', in the form of annuities, raises social welfare by $\Sigma\ (W_i/r_i)g_i'$. The appropriate weights to use in evaluating future gains are the "discounted" versions of those used to evaluate present gains. Conversely, if $\Sigma\ W_i' g_i'$ gives the social value of annuity package g_i', then $\Sigma\ W_i' r_i g_i$ gives the social value of present dollars net gain package g_i.

Thus, provided the individual discount rate are known, it makes no difference to the analysis whether the dictator expresses his evaluation of changes in distribution in terms of present dollars or future perpetuity dollars, since an analysis in terms of the one can be converted into an equivalent analysis in terms of the other. The test formulas which emerge in terms of weights, benefit and tax shares, and individual discount rates, of course, differ, depending upon the numeraire for which the weights are appropriate. Again recall that individual i's present value gain from a project is

$$\frac{b_i Y}{r_i} - t_i K,$$

and his equivalent perpetuity gain is

$$b_i Y - t_i r_i K$$

If present dollars are the numeraire in terms of which the effect of transfers is reckoned, and W_i the corresponding weight to be given to i's present dollar gains, then the correct weights for gains expressed in perpetuity terms are W_i/r_i. The extent of welfare gain from the project is obviously equivalently represented by the algebraically equivalent expressions,

$$\sum W_i\left(\frac{b_i Y}{r_i} - t_i K\right)$$

and

$$\sum \frac{W_i}{r_i}(b_i Y - t_i r_i K),$$

and the condition of positive welfare gain defines a critical rate of return,

$$\frac{Y}{K} \geqq \frac{\sum W_i t_i}{\sum \frac{W_i b_i}{r_i}} \tag{5}$$

A simple manipulation allows us to write this condition as

$$\frac{Y}{K} \geqq \frac{\sum W_i t_i}{\sum W_i t_i \left(\frac{b_i}{t_i r_i}\right)} \tag{6}$$

by which the critical rate is seen to be the harmonic mean of the quantities $t_i r_i/b_i$, with weights $W_i t_i$. For any individual, i, the quantity $(t_i/b_i)\ r_i$ gives the minimum rate of return which must be attained by a project with the given tax-benefit distribution before it yields a net benefit to him. It might be referred to as i's "cut-off rate" for such projects. The lower his tax share, the higher his benefit share, and the lower his rate of time preference, the lower is this minimum rate of return.

If future perpetuity dollars are the basis for the welfare calculation the critical rate is a different average of individual cut-off rates. Let W_i' be the weights attached to perpetuity gains, so that $r_i W_i'$ are the correct weights in present dollar terms. The condition of net social gain can be written,

$$\frac{Y}{K} \geqq \frac{\sum W_i' t_i r_i}{\sum W_i' b_i} \tag{7}$$

Again, a simple manipulation allows this condition to be written in the form,

$$\frac{Y}{K} \geq \frac{\sum W_i' b_i \left(\frac{t_i r_i}{b_i}\right)}{\sum W_i' b_i}, \tag{8}$$

by which the critical rate is seen to be the arithmetic mean of individual cut-off rates, with weights $W_i' b_i$.[18]

If the dictator is unrestricted in his ability to transfer present dollars, he will see to it that the weights W_i are all unity, and hence that the weights W_i' are equal to $1/r_i$. Substituting in either formula yields the present value test. If future transfers may be made freely, the weights W_i' will all be unity, and the weights W_i will be equal to r_i. Substituting in either formula yields the perpetuity value test.

[18] The individual cut-off rates would presumably also be the relevant variables to consider in any modelling of the political decision process. In his original presentation of the paradox as well as in correspondence Dorfman has suggested that in a one-man one-vote majority rule system the critical rate for social decision would be the median of individual discount rates. However this will be the case only if each individual's tax share equals his benefit share. More generally we should expect the critical rate for a project to be the median of individual cut-off rates, $t_i r_i / b_i$, when the decision is made by referendum. In exploring the implications of this the fact should be taken into account that the decisions which are referred directly to the voters generally concern whole programs or packages of programs, rather than isolated projects.

REFERENCES

K. J. Arrow, *Social Choice and Individual Values*, 2nd ed., New York 1963.

O. Eckstein, "A Survey of Public Expenditure Criteria," in Universities-National Bureau Committee for Economic Research, *Public Finances: Needs, Sources and Utilization*, Princeton 1961, pp. 453–463, 503–504.

J. Hirshleifer, "Comments" on Eckstein, in Universities-National Bureau Committee for Economic Research, *Public Finances: Needs, Sources and Utilization*, Princeton 1961, pp. 495–503.

J. V. Krutilla and O. Eckstein, *Multiple Purpose River Development*, Baltimore 1958.

R. C. Lind, "The Social Rate of Discount and the Optimal Rate of Investment: Further Comment," *Quart. J. Econ.*, May 1964, *78*, 336–45.

S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quart. J. Econ.*, Feb. 1963, *77*, 95–111.

P. A. Samuelson, "Evaluation of Real National Income," *Oxford Econ. Pap.*, Jan. 1950, *2*, 1–29.

Gordon Tullock, "The Social Rate of Discount and the Optimal Rate of Investment: Comment," *Quart. J. Econ.*, May 1964, *78*, 331–36.

B. A. Weisbrod, "Income Redistributive Effects and Benefit-Cost Analysis," in Samuel B. Chase, Jr., ed., *Problems in Public Expenditure Analysis*, Washington 1968.

The Demand For International Reserves

By MICHAEL G. KELLY*

The introduction by the International Monetary Fund of Special Drawing Rights (*SDRs*) as a supplement to existing international reserves raises the crucial question of their appropriate initial size and long-run rate of growth. If they become part of the working balances of member countries as the *IMF* envisages, *SDRs* will stand to produce a major influence on the functioning of the international monetary system. Differing volumes and growth rates of *SDRs* will have different impacts on the balance of payments of deficit and surplus countries and cause a major realignment of the burden of adjustment. It is desirable then to determine a policy for the creation of these rights which will produce an expansion of international reserves which is in some sense optimal.

An appropriate supply policy for reserves creation can only be specified if something is known about the demand for reserves. Without some knowledge of governments' demand, the effects of a change in supply cannot be predicted. Much of the literature in the postwar period has focussed on the adequacy of existing reserves.[1] The "adequacy" of or "need" for reserves, however, is not equivalent to the demand for reserves. The latter reflects the policy judgments of the reserves holders themselves and may or may not correspond to any objective assessment of adequacy. Only more recently has the more relevant question been raised, what determines the reserves that governments actually hold.[2]

Reserve holdings can be regarded as resulting from explicit policy decisions. The specification of alternative means of handling payments imbalances yields a demand function which depends on the relative benefits and costs of the alternatives. Section I of this paper presents a model yielding an optimal level of reserves for a government attempting to maintain external and internal balance under a regime of pegged exchange rates. The government maximizes utility subject to the trade-off between lower income levels implicit in large reserve holdings and greater income fluctuations generated by exogenous external disturbances which cannot be neutralized when reserve holdings are small. The model was tested on a group of 46 countries over the period 1953–65; the results are presented in Section II. The final section considers the implications of the study for international monetary reform.

I. *The Determinants of Reserve Holdings*

If it wishes to maintain a fixed external value for its currency, a government[3] must

* The author is assistant professor of economics at the University of British Columbia. This study is based on his doctoral dissertation submitted to the University of Chicago in 1968. Much of the work was carried out under a research fellowship from The Brookings Institution. Acknowledgement is made of helfpul criticisms of earlier versions given by Robert A. Mundell, Harry G. Johnson, Lester G. Telser, Sidney L. Carroll, and a referee.

[1] See, for example, James E. Meade (pp. 207–10), International Monetary Fund (1953, pp. 196–212); (1958, pp. 19–30), Robert Triffin (pp. 36–46), Oscar L. Altman (pp. 223–42), Thomas Balogh (pp. 344–45), and Tibor Scitovsky (pp. 101–09).

[2] Leland B. Yeager (pp. 348–49), Fritz Machlup (1964, pp. 264–76), Hang-Shang Cheng, Heinz Robert Heller (pp. 197–304), and Peter B. Clark. Empirical findings have been presented in Weir M. Brown (pp. 12–17), Peter B. Kenen and Elinor Yudin, Heller (pp. 304–9), Machlup (1966, pp. 178–99), and T. J. Courchene and G. M. Youssef.

[3] No distinction is made in what follows between the treasury, foreign exchange authority, and central bank. It is assumed that all three act in concert to achieve the joint objectives of the "government."

act to clear the market for foreign exchange at that desired value. This will require some action on its part except in coincidental cases where the quantities demanded and supplied by others at that price are equal.

The government must therefore adopt policies which cause demand and supply to shift or it must itself act as residual supplier, standing ready to offer unlimited quantities of foreign and domestic currencies at the pegged price. This latter method, of course, requires the government to possess sufficient quantities of both to satisfy all potential demand. Domestic currency poses no problem when the government itself commands the printing press. But foreign currencies are limited: the amount available, in fact, will equal the sum of historical net surpluses in the country's balance of payments.[4]

Since such surpluses represent the surrender of equivalent resources the authority is constrained in building up a stock of foreign currencies or reserves. In determining the reserves it wants to hold, a government will weigh their resource cost against the costs of alternative balance-of-payments policies intended to isolate the domestic economy from external disruptions.

Consider a government's response to a shift in the demand for foreign currency or supply of domestic currency, resulting in an excess demand[5] for foreign exchange at the pegged price. In the simplest case, the exchange authority can allow its reserves to run down to satisfy the excess demand. If it sterilizes the reserves loss, the domestic money supply will be unaffected and the domestic economy will not be subjected to the deflationary pressures inherent in following the gold standard "rules of the game." The more frequently the government does this, the greater will be the average balance of payments deficit and the larger the necessary average level of reserves. Alternatively the government can adopt policies which will allow adjustment, causing the deficit to be reduced or eliminated. By permitting monetary contraction to exert deflationary pressure, the government can eliminate the private sector's excess demand for foreign exchange. This is essentially the gold standard method and requires domestic income and price levels to vary in response to balance of payments changes.

This method is often rejected precisely because of its deflationary discipline, which conflicts with the goals of domestic stability. In its stead, a government may resort to commercial policy, which can be defined broadly to include all measures designed to alter the choices of residents and nonresidents for domestic versus foreign commodities and debt instruments.[6] Commercial policy allows the foreign exchange market to be equilibrated with smaller reserve holdings on average and with less disruptive effects on the domestic economy. The gain is illusory however, since the resource misallocation due to the commercial policy offsets the savings of reserves.[7] Stabilizing the level of real income still requires larger reserves. The final option is to alter the peg at which the foreign exchange rate is supported. Such a policy would also require the holding of less reserves on average but would expose the economy to the same price fluctuations as under the gold standard method.

[4] Defined to include domestically mined gold obtained by the authority.

[5] For simplicity of exposition the following analysis will be confined largely to balance-of-payments deficits but it is equally valid, *mutatis mutandis*, for balance-of-payments surpluses.

[6] The distinction between deflationary and commercial policies is implicit in Meade (pp. 48–49; 252–53). Alternative dichotomies can be found in Harry G. Johnson (pp. 162–68), Gottfried Haberler (p. 29), and Machlup (1965, pp. 188–208).

[7] It is assumed that any welfare maximizing optimum tariffs will be set independently of the current status of the balance of payments.

If the shift in demand for foreign exchange is temporary rather than permanent, elimination of the direct connection between the balance of payments and the domestic economy may be particularly desirable. For example, a deficit may be due to a cyclical decrease in foreign demand which is expected to increase again at a later date. The offsetting of reserve changes under these conditions will isolate the domestic economy from foreign fluctuations and obviate cyclical deflation and and inflation otherwise required to accommodate balance-of-payments changes without the use of commercial policies or exchange rate changes. To fulfill this stabilization role, reserves must be large enough to absorb exogenous cyclical swings in the balance of payments. Given the probability distribution of such swings and some fixed risk of reserves depletion there exists a set trade-off between average reserve holdings and income variability.

The cost of reserves is the income sacrificed in holding them, the alternative return which the assets could yield. This is the social opportunity cost of capital, less any yield obtained on the reserve holdings themselves. The higher this capital cost the greater the incentive to economize on reserve holdings and to allow greater fluctuations in income.

Reserve holdings will thus be determined by a government according to the contribution they make to its stabilization goal, the costs they entail, and the alternative means available to that government to handle balance-of-payments deficits. Optimum reserves will equate at the margin the value attached to reduced instability of income and the value of the reduction in income such reserves entail.

We can derive the optimum level of reserves as a function of the factors outlined above, in explicit form, once the exact relationships among the variables are specified. The following model is intended to demonstrate the procedure using some plausible but highly restrictive assumptions.

Consider a country which maintains a pegged value for its currency by using a buffer stock of gold and foreign currencies. This stock of reserves, R, has been accumulated through past balance of payments surpluses and is held for stabilization purposes. The only source of variations in the income level is changes in exports, X, assumed to be exogenously determined.

Starting from an initial position of equilibrium the change in reserves in any time period will equal exogenous changes in foreign demand less any induced changes in domestic demand for foreign commodities.

$$\Delta R_t = \Delta X_t - \Delta M_t \tag{1}$$

where M represents imports of goods and services.[8] Imports are endogenous and subject to the influence of government policy. In general the government will permit or induce changes in imports to offset partially the change in exports. This will be true whether the government responds to the exogenous change by allowing the money supply to follow *pari passu* or by altering tariffs or the exchange rate. Thus we can define an import response coefficient, f, equal to dM/dX. This is a policy variable and reflects the government's willingness to allow the external disturbance to spill over to the domestic economy. It will normally lie between zero and one. Related to f will be an income response coefficient, g, equal to dY/dX, which reflects the effect of the change in exports on the level of income. Variables f and g are positively correlated: any induced change in imports will accompany

[8] In principle, net capital imports can be treated as endogenous and added to imports of goods and services to obtain the total induced changes. Capital flows, however, are ignored in the remainder of this section.

a change in real income in the same direction. For simplicity we will assume the relationship is linear homogeneous so that $f = mg$ where m is now the marginal propensity to import.[9] We can then calculate the variance in both reserves and the level of income as a function of the one policy variable and the variance of external disturbances, on the further assumption that the country is small enough to avoid producing feedback effects. Thus

$$(2) \qquad \Delta R = \Delta X(1 - f),$$

$$(3) \qquad V(R) = E(\Delta R^2) = V(X)(1 - f)^2,$$

$$(4) \qquad V(Y) = E(\Delta Y^2) = g^2 V(X),$$

where V denotes variance and E the expectation, with changes measured from the mean. The larger are f and g, the less will reserve levels vary and the greater will be the fluctuations in the level of income.

Consider now that the government wishes to avoid running out of reserves, or more realistically that there is some minimum level of reserves, R', below which it becomes prohibitively costly to pursue stabilization policies. The government then will maintain average reserves sufficiently large so that the probability of actual reserves falling below the fixed target level is small.

$$(5) \qquad P[R < R' \mid E(R), V(R)] = e$$

where e is some small number. Then given the probability distribution of exogenous balance-of-payments changes, ΔX, we can determine exactly the average level of reserves, $E(R)$, needed to maintain e. For any regularly behaved probability density function we can say that, given e, $dE(R)/dV(R) > 0$. The larger the variance in reserves, the larger their average must be to maintain a given probability that they will not fall below a specified level. Let us assume that this probability varies directly with $V(R)$ and inversely with the square of $E(R)$.[10] The form

$$(6) \qquad e = cV(R)/E(R)^2 \qquad c > 0$$

has the properties $\partial e/\partial E(R) < 0$, $\partial^2 e/\partial E(R)^2 > 0$, $\partial e/\partial V(R) > 0$, and, given e, $dE(R)/dV(R) > 0$. Then, from equations (3) and (6), the average level of reserves will be:

$$(7) \qquad \begin{aligned} E(R) &= (c/e)^{1/2} S(R) \\ &= (c/e)^{1/2} S(X)(1 - f) \end{aligned}$$

where S denotes the standard deviation.

Thus the average level of reserves will be determined by their standard deviation which in turn is a function of the balance of payments policy variable. There is a trade-off between the level of reserves and the variance of income: the smaller f, the larger the former and the smaller the latter. Since larger reserves reduce income, the government must choose between a higher income level with large fluctuations about that level and a lower income level with smaller fluctuations. From equation (4) we have $g = S(Y)/S(X)$. Setting $f = mg$ and substituting into equation (7), we obtain the technical relationship between average reserve holdings and income variability:

$$(8) \qquad E(R) = (c/e)^{1/2}[S(X) - mS(Y)]$$

If we introduce a utility function expressing the country's preferences for a lower level of income versus reduced instability of income, we can determine its optimum reserves holdings. The reduction

[9] m is determined by the choice of f and thus is not the conventional autonomous marginal propensity.

[10] The square-root rule and similar inventory model theories suggest that the marginal probability of falling below (R') decreases at an increasing rate as we reduce (R') by equal decrements when the exogenous changes (ΔX), are generated by a random process.

The relationship between $V(R)$ and $E(R)$ depends on the actual distribution of reserve changes and the particular point of the distribution at which e is chosen. Using Chebychev's Inequality, Peter B. Clark (pp. 17–18) has developed the following relationship, assuming a symmetrical distribution and $R' = 0$:

$$e < V(R)/2E(R)^2$$

in income due to tying up resources in reserves will be

$$(9)\qquad Y' - Y = Ri$$

where Y' is total output available if no reserves are held and Y output available after providing for reserves at their opportunity cost, i. Consider the following utility function:

$$(10)\qquad \begin{aligned} U = &- a[E(Y') - E(Y)]^2 \\ &- b[(Y) - E(Y)]^2 \qquad a, b > 0 \end{aligned}$$

This expresses utility as a negative quadratic function of the reduction in the average level of real income and of fluctuations about that level. This gives us the desirable properties of increasing marginal disutility from reduced income and from increased income variability. Then by substitution from (9) we get expected utility as

$$(11)\qquad E(U) = - ai^2E(R)^2 - bV(Y)$$

Maximizing (11) over $E(R)$ and $S(Y)$ subject to (8) we obtain the optimum level of average reserves in terms of our structural parameters.

$$(12)\qquad \widehat{E(R)} = \frac{S(X)}{(e/c)^{1/2} + (c/e)^{1/2}(m)^2i^2(a/b)}$$

The optimum level of reserves varies directly with $S(X)$, the standard deviation of exogenous stocks, and b, the marginal disutility of income variance. It varies inversely with a, the marginal disutility of income reductions, with i, the cost of reserves, and with m, the marginal propensity to import. Optimum reserves will vary inversely with e, the risk of reserves dropping below some specified level, unless a is greater than b by a very large factor.

The relationship between reserves and income stabilization can be illustrated by Figure 1. From equations (3), (4), (6), and (8) we have:

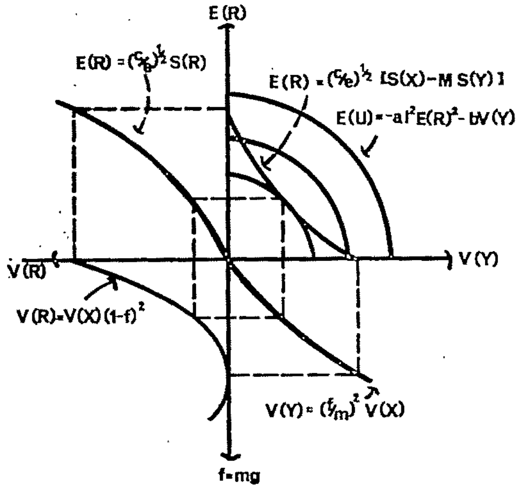


FIGURE 1

$$\frac{\partial E(R)}{\partial V(R)} > 0 \qquad\qquad \frac{\partial^2 E(R)}{\partial V(R)^2} < 0$$

$$\frac{\partial V(R)}{\partial f} < 0 \quad \text{for } f < 1 \quad \frac{\partial^2 V(R)}{\partial f^2} > 0$$

$$\frac{\partial f}{\partial V(Y)} > 0 \qquad\qquad \frac{\partial^2 f}{\partial V(Y)^2} < 0$$

$$\frac{\partial E(R)}{\partial V(Y)} < 0 \qquad\qquad \frac{\partial^2 E(R)}{\partial V(Y)^2} > 0$$

The transformation curve between average reserves and the variance of income is thus convex. When $E(R) = 0$, the variance of income is at a maximum. The balance of payments is adjusted by allowing changes in imports to offset completely changes in exports: f equals one. As f decreases, more of the exogenous external disturbances are sterilized and reserves must accordingly be larger. When $V(Y) = 0$, average reserves are at a maximum. The domestic economy is completely isolated from the foreign sector since f equals zero. The figure also illustrates why f will normally be less than one. For larger values both $E(R)$ and $V(Y)$ increase. By "over-reacting" to exogenous changes the government can make them destabilizing. The transformation curve

is a lower bound for $E(R)$ and $V(Y)$ combinations.

Our utility function has $\partial E(R)/\partial V(Y) > 0$ and $\partial^2 E(R)/\partial V(Y)^2 < 0$ and generates a family of concave indifference curves. Since both $E(R)$ and $V(Y)$ are "bads," increasing utility is achieved by moving towards the origin. The shapes of the curves thus assure the attainability of a unique equilibrium.[11]

II. *Empirical Evidence on Reserve Holdings*

The model presented above yields an optimum average level of reserves as a function of exogenous changes in the balance of payments, the marginal propensity to import, the opportunity cost of capital, and the government's preferences between income levels and income variability. If such preferences are similar across countries then differences in average reserve holdings can be attributed to the remaining three variables.

To test the model, annual data were gathered for 46 countries for the period 1953–65 (going back to 1949 for some balance of payments items). Reserves were measured by the *IMF* definition and averaged over a period of one to six years, ending in the current year, the period varying inversely with a country's average propensity to import.[12]

Exogenous changes in the balance of payments are measured by the standard deviation of exports of goods and services over the preceeding four years and the current year. While exports are not completely exogenous and not all imports endogenous, this dichotomy permits straightforward measurement and is more plausible than any simple alternative.

Initial calculations of the observed "marginal propensity to import" revealed wide variations over the time period examined rather than the constant predicated in our model.[13] This phenomenon presumably reflects the interference of "commercial policy" violating the *ceteris paribus* assumption implicit in the marginal propensity to import concept.[14] The current year's average propensity to import, which is much more stable, was accordingly substituted for the marginal propensity; this will be a biased measure to the extent that the income elasticity of demand for imports varies across countries.

The opportunity cost of reserves is the most difficult variable to handle. It is usually measured by some rate of interest, such as the yield on long-term government bonds. However, data on any rates of interest are not available for many of the countries here considered. The poor or nonexistent capital markets of most of the less developed countries provide little

[11] We can look at the first quadrant as the obverse of the usual utility-transformation curves map. Consider the curves as emanating from an origin diagonally across from the one drawn. Then the axes measure "goods" (income level and income stability) and the curves have their usual shapes and interpretations.

[12] The use of different averaging periods is made on the hypothesis that less open countries (those whose foreign trade is relatively small) can afford to allow greater deviations from the desired average than more open countries can. The appropriate period is therefore shorter for the latter group than for the former. The six-year grouping is an arbitrary one dictated by the availability of data and not by any a priori reasoning. Several of the regressions described in the text were rerun using the following alternative definitions of reserves: gold plus foreign exchange holdings; current holdings on IMF definition; and IMF-defined reserves averaged over five years for all countries. The difference in the multiple correlation coefficients and the standard errors of the independent variables was very small, with the five-year average definition usually performing slightly better and the other two slightly poorer than the definition used in the text.

[13] If net capital flows are treated as endogenous and included in the measured marginal propensity the variations are substantially greater.

[14] Half of the countries studied have changes in imports more highly correlated with changes in exports than changes in income for the 1953–65 period, suggesting that imports are in fact "controlled" by balance of payments considerations. This would indicate that f is the appropriate variable to analyze; if g is constant, however, the two are equivalent.

indication of the scarcity of resources relative to the more developed countries. What little evidence there is suggests that international interest rate differentials are not substantial in any case.[15] However, to explore the possibility that opportunity costs are in fact higher in less developed countries, two alternative measures of capital scarcity were tried separately as independent variables.

The first is the level of per capita income. A priori we would expect that the poorest countries might have the greatest need of capital and the highest opportunity costs. The inclusion of per capita income has the added advantage of indicating any wealth effects on reserve holdings. We might expect the richer countries to hold greater reserves even if their opportunity costs were no lower.

The second measure of opportunity cost is the balance of international indebtedness of a country. Whereas this reflects past relative capital scarcities, as opposed to current scarcities, it may provide a long-run view of alternative uses of reserves. It, too, contains a wealth effect element: countries with large foreign assets are presumably better off than countries with large foreign debts.[16] Data on international indebtedness are not available but data on interest and dividend payments are. These were used to represent the corresponding assets and liabilities.[17]

Income levels are measured by gross national products expressed in local currencies, converted at existing exchange rates. This is the most appropriate method of deflating nominal values if the exchange rate is an equilibrium one: to the extent that internal inflation leads to devaluation only after a lag, the real foreign equivalent of domestic value so converted is, of course, overstated. In principle, however, there is no need for further adjustment of these foreign currency figures. All values are expressed in nominal *U.S.* dollars; these were not deflated.

The level of reserves, R^*, averaged as described above, was regressed on the standard deviation of exports, $S(X)$, the average propensity to import, M/Y, and the two alternative proxies for opportunity cost: per capita income, Y/P, and foreign assets and liabilities, A and L in both linear and logarithmic forms. The pooled data contain 598 observations (46 countries$\times$13 years). Dummy variables were included for each of the 46 countries to test for inter-country differences. The *log* forms yielded superior fits. The resulting equations (with dummies are not shown) are:

$$\text{(13)} \quad \log R^* = \underset{(3.9)}{.117} \log S(X) + \underset{(7.6)}{.777} \log M/Y + \underset{(13.0)}{.904} \log Y/P \qquad R^2 = .968$$

$$\text{(14)} \quad \log R^* = \underset{(6.4)}{.194} \log S(X) + \underset{(2.8)}{.306} \log M/Y + \underset{(7.6)}{.112} \log A + \underset{(3.3)}{.062} \log L \qquad R^2 = .964$$

The *t*-ratios are shown in parentheses below each coefficient. All of the independent variables are significant[18] and all but the import propensity and foreign liabilities

[15] See U. Tun Wai (1956, 1957).

[16] Although large foreign assets can also be viewed as a substitute for reserve holdings and may thus reduce optimum levels of the latter.

[17] The representation is exact only if average rates of return and the proportion of retained earnings to total earnings are everywhere the same.

[18] All tests of significance cited are at the 95 percent level, unless otherwise specified.

variables have the expected sign. The latter's positive value suggests that it captures a precautionary-balances effect rather than a capital scarcity effect. The positive coefficient for the average propensity to import may have two explanations: it is not representative of the marginal propensity, or the marginal propensity has a positive influence on reserves because there are income fluctuations due to internal exogenous shifts in demand as well as external shifts.

The dummy variables for each country, representing the average reserve holdings unexplained by the independent variables, are in most cases significantly different from the group mean, indicating that the independent variables alone do not account for all of the intercountry differences in reserve holdings. The use of dummy variables allows a test of whether countries hold reserves that are on average "high" or "low" when account is taken of the variability of their export earnings, their propensity to import, and their per capita incomes or balance of indebtedness. However, the construction of equations (13) and (14) implicitly assumes that all countries respond in the same way to changes in these variables at the margin. A test for differences in the coefficients of the exogenous variables would require separate regressions for each of the countries. The presence of serial correlation and the small range of variation of the variables over a 13-year period make analysis of such regressions of doubtful value. To obtain some measure of international differences in the coefficients the countries were divided into two equal groups by two different criteria. First, the countries were split into more developed and less developed (ranked by per capita income) and, second, they were split into more open and less open (ranked by the ratio of imports to income). Regressions were run on the exogenous variables for each of these without country dummies. Separate regressions were run for all countries together with one dummy to capture differences on average for the group.

The resulting regressions are shown in Table 1. The most noticeable changes are in the signs of the propensity to import and per capita income variables which become negative, indicating strong intracountry negative correlations with reserve levels. The negative coefficient of the averare propensity to import when intra-country differences enter the equation does indicate that reserve levels fall as imports rise: if imports are controllable this pattern confirms the adjustment role played by the marginal propensity to import in our model. However this pattern is at variance with the predicated correlation between exports and imports. The change in sign also reflects the variability in the measured marginal propensity to import and confirms the difficulty of capturing the true relationship between exports, imports, and real income.[19]

The negative value of the per capita income variable suggests that, while it may have some relevance as a long-run indicator of opportunity cost, in the short run it is too highly correlated with other variables to offer additional information. The exports variance variable is substantially higher than in the initial equations where country dummies are included. The inclusion of dummy variables for the more open and more developed countries in the grouped equations suggest there is little difference between the groups on average when the explanatory variables are taken into account. F-tests on the sum of squared residuals of the separate regressions versus the sum of squares of the pooled data lead to rejection of the hypothesis that the coefficients are the same for each group however divided. (The hypothesis is ac-

[19] See the discussion at the beginning of this section.

TABLE 1—COEFFICIENTS OF INDEPENDENT VARIABLES FOR $\log R$*

	Constant	$\log S(X)$	$\log M/Y$	$\log Y/P$	$\log A$	$\log L$	Dummy	R^2
Less Open Countries	1.746	.941	−.453	−.180				.847
		(28.6)	(4.2)	(3.6)				
	2.244	.552	−.363		.191	.034		.897
		(15.5)	(4.1)		(12.6)	(1.8)		
More Open Countries	1.348	1.037	−.165	−.090				.800
		(28.3)	(1.0)	(1.4)				
	1.913	.547	−.335		.220	.089		.864
		(11.5)	(2.4)		(11.3)	(3.1)		
All Countries	1.670	.993	−.290	−.152			.127	.821
		(40.7)	(3.2)	(3.8)			(1.3)	
	2.335	.552	−.274		.207	.053	−.213	.879
		(19.0)	(3.7)		(17.2)	(3.5)	(2.8)	
Less Developed Countries	3.500	.972	−.261	−.471				.690
		(27.1)	(2.7)	(6.2)				
	1.516	.702	−.518		.161	.002		.777
		(14.2)	(6.2)		(9.6)	(.0)		
More Developed Countries	−.910	.987	−.079	.279				.803
		(29.4)	(1.0)	(3.3)				
	2.001	.461	−.300		.231	.176		.900
		(11.8)	(5.3)		(13.6)	(6.6)		
All Countries	1.950	.988	−.203	−.168			.082	.821
		(38.8)	(3.3)	(2.9)			(.5)	
	1.973	.594	−.394		.202	.047	−.171	.879
		(18.4)	(7.9)		(16.9)	(3.0)	(2.6)	

The t-ratios are shown below each coefficient.

ceptable at the 99 percent level when countries are divided by openness and assets and liabilities are included among the independent variables.)

The analysis of the previous section also suggests that average reserve levels will be independent of the alternative balance-of-payments policy tools used, or rather that their joint use will lead to the same reserves policy in all cases. Since the other policy options are difficult to quantify their use was "measured" qualitatively and indicated by three dummy variables as follows: D_1, adherence to a pegged exchange rate regime, equals one if the country is committed to maintaining the existing par value of its currency;[20] D_2, use of commercial policy for balance of payments purposes, equals one if a country has declared its currency "convertible" under Article VIII[21] of the *IMF* Articles of Agreement; D_3, use of monetary policy for balance of payments adjustment, equals one if a

[20] This index was based largely on an article by M. G. De Vries.

[21] This Article provides that a member undertake not to impose restrictions on trade, to avoid discriminatory currency arrangements, and to redeem its own currency for gold when tendered by other members. Exceptions are made for the redemption of currency not acquired through current trade. Thus convertibility status does not extend to all transactions; even then countries which have declared their currencies convertible have subsequently resorted to restrictive trade measures for balance-of-payments purposes. However, convertibility status is a convenient, if very imperfect, measure of a country's use of commercial policy in handling balance-of-payments adjustment. Experimentation with a six-level index based on reported balance-of-payments restrictions added little extra explanatory power. See p. 60 of my doctoral thesis.

country's money supply has *not* shown any evidence of responding to changes in reserves, i.e. the rules of the gold standard game are not being followed.[22] All three coefficients should be positive; maintenance of a pegged exchange rate, abstention from commercial policy, and pursuit of monetary policy independently of the balance of payments will all require large reserve holdings on the average. The initial equations were rerun with the new policy dummies and a new equation was run with only the policy dummies and the 46 country dummies. The results (with country dummies not shown) are:

$$(15)\quad \log R^* = \underset{(3.3)}{.100} \log S(X) + \underset{(6.6)}{.682} \log M/Y + \underset{(9.5)}{.762} \log Y/P + \underset{(5.4)}{.352} D_1 + \underset{(2.3)}{.143} D_2 + \underset{(.4)}{.041} D_3 \qquad R^2 = .970$$

$$(16)\quad \log R^* = \underset{(4.8)}{.144} \log S(X) + \underset{(2.6)}{.271} \log M/Y + \underset{(5.9)}{.085} \log A + \underset{(2.6)}{.045} \log L + \underset{(4.7)}{.324} D_1 + \underset{(6.0)}{.347} D_2 + \underset{(1.0)}{.112} D_3 \qquad R^2 = .968$$

$$(17)\quad \log R^* = \underset{(7.5)}{.537} D_1 + \underset{(9.8)}{.584} D_2 + \underset{(.6)}{.065} D_3 \qquad R^2 = .961$$

All of the independent variables remain significant. Two of the three policy dummy variables are also significant with the expected sign. This is also true in equation (17) where all of the other independent variables are dropped.[23] The particular monetary policy followed, at least as captured by our crude method, does not seem to influence reserve levels but adherence to exchange rate parity and formal abstention from trade interference evidently to lead to higher reserves. The latter three are a policy "package" and tend to be adopted simultaneously.

Finally, to test for the stability of reserve holdings over time, separate cross-sectional regressions were run for each of the 13 years. Dummy variables were added to the pooled data for each of the years, 1954–65. The results for the second set are shown in Table 2.[24] An F-test on the squared residuals leads to acceptance of the hypothesis that the coefficients have not changed over the 13-year period under either measure of opportunity cost.

Export variability is consistently the strongest explanatory variable. The propensity to import variable has a negative sign throughout as predicted, increasing somewhat our confidence in the role of the marginal propensity. The dummy variables in the pooled data show a slight trend peaking in 1959; the average level falls below that in 1953 for the first time in 1965.

The tests of significance cited above generally presume that the error terms of the equation are independently distributed. However, examination of the residuals of equations (13) and (14) reveals serial correlation exists in many cases. A

[22] This was determined by examination of the annual changes in reserves and the money supply over the entire period. In a few cases a change in policy seemed to occur and the index was shifted accordingly at the point of change.

[23] The high R^2 in equation (17) is due to the inclusion of dummy variables for each of the countries. Running the equations without the latter yields an R^2 of only .122.

[24] The general pattern of the first set is the same; the coefficient of the per capita income variable is negative in all years.

TABLE 2—COEFFICIENTS OF ANNUAL CROSS-SECTIONS FOR $\log R^*$

Year	Constant	$\log S(X)$	$\log M/Y$	$\log A$	$\log L$	R^2
1953	1.786	.583 (4.5)	−.452 (2.2)	.203 (3.9)	.011 (.0)	.821
1954	2.160	.504 (4.0)	−.483 (3.0)	.212 (5.0)	.035 (.7)	.871
1955	2.316	.540 (3.7)	−.365 (2.0)	.207 (4.0)	.028 (.5)	.834
1956	1.990	.571 (5.9)	−.524 (3.1)	.201 (5.0)	.003 (.0)	.889
1957	2.071	.464 (4.6)	−.654 (3.6)	.207 (5.4)	.055 (.8)	.880
1958	2.020	.567 (6.3)	−.536 (3.3)	.180 (4.9)	.014 (.3)	.901
1959	1.814	.643 (5.9)	−.519 (3.0)	.182 (4.1)	.002 (.0)	.888
1960	2.067	.552 (5.7)	−.418 (2.7)	.224 (6.0)	.052 (1.1)	.925
1961	1.898	.618 (5.8)	−.392 (2.3)	.192 (4.2)	.037 (.6)	.910
1962	1.693	.625 (5.2)	−.371 (2.0)	.182 (3.6)	.077 (1.3)	.902
1963	1.677	.566 (4.1)	−.247 (1.3)	.156 (2.7)	.211 (2.9)	.903
1964	1.629	.599 (4.6)	−.140 (.8)	.167 (3.3)	.185 (2.2)	.897
1965	1.690	.516 (4.2)	−.215 (1.2)	.217 (4.2)	.182 (2.3)	.892
1953–65[a]	1.770	.578 (19.2)	−.420 (8.8)	.196 (16.0)	.046 (2.9)	.882

The t-ratios are shown below each coefficient.

[a] For the pooled regression, dummy variables were added to represent each individual year. The coefficients and their t-values are shown below.

$$\underset{(1.1)}{+.144\,T_{54}}\underset{(2.1)}{+.256\,T_{55}}\underset{(1.8)}{+.218\,T_{56}}\underset{(1.5)}{+.193\,T_{57}}\underset{(2.0)}{+.250\,T_{58}}\underset{(2.3)}{+.287\,T_{59}}\underset{(2.1)}{+.261\,T_{60}}$$

$$\underset{(1.7)}{+.209\,T_{61}}\underset{(1.1)}{+.137\,T_{62}}\underset{(1.1)}{+.134\,T_{63}}\underset{(.0)}{+.024\,T_{64}}\underset{(.6)}{-.080\,T_{65}}$$

simple runs test on the residuals of the pooled data for each of the 46 countries separately indicates that roughly half have significantly positive autocorrelations. Many of these are European and Asian countries which built up their reserves in the 1950's after wartime and reconstruction period depletion; on the

other hand, a few Latin American and Asian countries ran down their large reserves accumulated during the war. There were no instances of negative correlation. This suggests that the averaging period used in constructing R^* is too short to capture long-run reserve levels; a longer period might reduce serial correlation substantially.[25]

Apart from the specification and measurement of the relevant variables the neglect of dynamic factors of adjustment is probably the major element in the unexplained variance of reserve levels. The essentially static nature of the model ignores the problem of how quickly a government recognizes secular changes in its balance of payments and how it adjusts to a new optimum level of reserves. Since such recognition presumably occurs only after cumulative imbalances, adjustment involves two parts: restoration of the old average and change to a new level. Appropriate specification of the process by which this occurs would be essential for any short-run implications to be drawn from the analysis.

III. *Implications for International Monetary Reform*

The recasting of the problems of the present international monetary system in terms of the demand for reserves directs attention to one aspect which is frequently ignored or whose importance is underestimated in most discussions. The empirical evidence presented here lends some support to the idea that reserve holdings are rationally determined and are related to a small set of variables common to all countries over a period sufficiently long to encompass major cyclical swings.

In particular the constancy of the coefficients for the annual cross-sections over a 13-year period provides a strong base on which to predict the growth in demand for reserves. If export variance and foreign assets and liabilities[26] continue to grow at the same rates (5.8, 10.4, and 9.8 percent respectively) and the same import-responsiveness is maintained, then the demand for reserves will increase at an average annual rate of 5.9 percent. The *SDR*'s to be created over the next three years will augment international reserves at a rate of just under 5 percent, leaving a small deficiency to be met by increased *IMF* quotas or additions to foreign exchange holdings through deficits in the balances of payments of the key-currency countries.[27]

Such a growth rate would be neutral, leaving the existing distribution of the international "burden of adjustment," restrictions on international trade, and deviations from parity exchange rates unchanged. A slower growth rate would increase and a faster rate decrease world inflationary pressures and the need to resort to such measures.[28]

[25] Cf. fn 12.

[26] It should be recalled that these are represented by interest and dividend payments. To the extent that assets and liabilities are the appropriate measures of opportunity costs changes in yields must be taken into account.

[27] That the average growth rate of total reserves in the 1953–65 period was less than 3 percent is due to such deficits, particularly that of the United States. This source of increases in reserves is of course limited and cannot act as a substitute for "international" reserves in the long run under current institutional arrangements.

[28] To the extent that trade restrictions impose negative externalities on the international monetary system the appropriate supply policy calls for expansion of reserves at a rate greater than the growth in demand; this would lead countries to hold larger reserves and conform to the *IMF* "rules" unless they were willing to permit the increased foreign demand to have a larger impact on their domestic economies. But this result depends on the existence of a type of "money illusion" on the part of reserves holders. Their willingness to accumulate more reserves on a continuing basis contradicts our earlier utility-maximizing assumptions. The attempt of all countries to spend the excess supply would generate world-wide inflation until the restoration of a new equilibrium with real values unchanged. Thus a policy of deliberate reserve creation to remove impediments to trade would be ineffective in the long run.

Finally it is interesting to note that the rate of growth in demand for reserves here derived is considerably lower than that which is implicit in the often used criterion of growth in world trade. While exports and imports have expanded at annual rates of 7.9 and 7.6 percent, respectively, the standard deviation of exports, which is the main determinant of the demand for reserves in our analysis, has grown at an average rate of only 5.8 percent. Thus a reduction in export variance in the 1970's would reduce the appropriate growth rate of reserves despite a growing volume of international trade. Similarly we can compute the income-elasticity of demand for international reserves: a continuation of the average annual increase in *GNP* of 6.3 percent yields an implicit elasticity slightly below unity.

REFERENCES

O. L. Altman, "Professor Triffin, International Liquidity, and the International Monetary Fund," in S. E. Harris, ed., *The Dollar in Crisis*, New York 1961.

T. Balogh, "International Reserves and Liquidity," *Econ. J.*, June 1960, *70*, 357–77; reprinted in H. G. Grubel, ed., *World Monetary Reform: Plans and Issues*, Stanford 1963.

W. M. Brown, *The External Liquidity of an Advanced Country*, Princeton Studies in International Finance No. 14, Princeton 1964.

H. S. Cheng, "A Theory of the Optimal Amount of Foreign Reserves of a Central Bank," unpublished doctoral dissertation, Princeton Univ. 1963.

P. B. Clark, "Optimum International Reserves and the Speed of Adjustment," mimeo. 1967.

T. J. Courchene and G. M. Youssef, "The Demand for International Reserves," *J. Polit. Econ.*, Aug. 1967, *75*, 404–13.

M. G. De Vries, "Fund Members Adherence to the Par Value Regime: Empirical Evidence," *Int. Monet. Fund Staff Pap.*, Nov. 1966, *13*, 504–30.

G. Haberler, *Money in the International Economy*, Cambridge, Mass. 1965.

H. R. Heller, "Optimal International Reserves," *Econ. J.*, June 1966, *76*, 296–311.

H. G. Johnson, "Towards A General Theory of the Balance of Payments," *International Trade and Economic Growth*, London 1958.

M. G. Kelly, "The Demand for International Reserves," unpublished doctoral dissertation, Univ. Chicago 1968.

P. B. Kenen and E. Yudin, "The Demand for International Reserves," *Rev. Econ. Statist.*, Aug. 1965, *47*, 242–50.

F. Machlup, *International Payments, Debts and Gold*, New York 1964.

———, "Real Adjustment, Compensatory Corrections, and Foreign Financing of Imbalances in International Payments," in Richard E. Caves, et al., eds., *Trade, Growth, and the Balance of Payments*, Chicago 1965.

———, "The Need for Monetary Reserves," Banca Nazionale del Lavoro, *Quart. Rev.*, Sept. 1966, *47*, 175–222.

J. E. Meade, *The Theory of International Economic Policy*, Vol. 1; *The Balance of Payments*, London 1951.

T. Scitovsky, *Economic Theory and Western European Integration*, Stanford 1958.

R. Triffin, *Gold and the Dollar Crisis*, rev. ed., New Haven 1961.

U. T. Wai, "Interest Rates in the Organized Money Markets of Underdeveloped Countries," *Int. Monet. Fund Staff Pap.*, Aug. 1956, *5*, 249–78.

———, "Interest Rates Outside the Organized Money Markets of Underdeveloped Countries," *Int. Monet. Fund Staff Pap.*, Nov. 1957, *6*, 80–125.

L. B. Yeager, "The Misconceived Problem of International Liquidity," *J. Finance*, Sept. 1959, *14*, 347–60.

International Monetary Fund, "The Adequacy of Monetary Reserves," *Staff Papers*, Oct. 1953, *3*, 181–227.

———, *Balance of Payments Yearbook*.

———, *International Financial Statistics*.

———, *International Reserves and Liquidity*, 1958.

International Diversification of Investment Portfolios

By HAIM LEVY AND MARSHALL SARNAT*

The theoretical models of portfolio selection developed by Harry Markowitz and James Tobin provide a positive explanation and normative rules for the diversification of risky assets, but the degree to which diversification can reduce risk depends upon the correlations among security returns. If the returns are not correlated, diversification could eliminate risk. Portfolio selection would become an analogue to the cancelling of risks in the insurance industry. On the other hand, if security returns are perfectly correlated, no amount of diversification can affect risk.

Within an economy, a strong tendency usually exists for economic phenomena to move more or less in unison giving rise to periods of relatively high or low general economic activity. As can be expected, this also holds true for individual securities and industries. For example, during the period 1951–67, a price index of *U.S.* industrial common stocks was positively correlated with indices of railroad and public utilities stock prices;[1] the correlation coefficients were 0.46 and 0.59, respectively.[2] The practical, as well as theoretical importance of portfolio selection stems from the fact that while the observed security returns for any particular country are highly correlated, they are not perfectly correlated, which implies the reduction (but not the elimination) of risk through diversification (see Markowitz, 1959 p. 5).

The existence of a relatively high degree of positive correlation within an economy suggests the possibility that risk reduction might be facilitated by diversifying securities portfolios internationally. This paper presents estimates of the potential gains from such diversification for the period 1951–67. A method for the empirical determination of the composition of optimal international portfolios is also presented, and some of the implications of international risk diversification for investment decisions and for the theory of international capital movements are explored.

I

In order to examine the potential gains (if any) accruing from international diversification, mean rates of return on common stocks and their standard deviations were calculated for 28 countries for the period 1951 to 1967 (see Table 1). The annual rate of return for each country was defined as the percentage change in the dollar value of its index of common stocks.[3]

$$(1)\qquad r_i(t) = \frac{P_i(t) - P_i(t-1)}{P_i(t-1)}$$

where:

* The authors are, respectively, visiting associate professor at the University of Illinois and senior lecturer at the Hebrew University, Jerusalem. They wish to acknowledge the helpful comments of Moshe Ben-Horim, Gad Marx, and a referee on an earlier draft of this paper. Part of the research was supported by grants from the Faculty of Social Sciences, Hebrew University and the Graduate Research Board, University of Illinois.

[1] The indices used were taken from various issues of the Federal Reserve Bulletin.

[2] The coefficient of correlation between railroad and public utility stock prices was 0.10 for the same time period.

[3] The indices used were obtained from various editions of *International Financial Statistics*. All calculations are on a pre-tax basis.

TABLE 1—MEAN RATES OF RETURN AND STANDARD DEVIATIONS OF COMMON STOCKS FOR 28 SELECTED COUNTRIES, 1951–67

(in percent)

Country	Rate of Return	Standard Deviation	Country	Rate of Return	Standard Deviation
Australia	1.2	18.2	Mexico	2.9	16.1
Austria	16.1	27.7	Netherlands	9.1	20.1
Belgium	3.2	10.7	New Zealand	4.4	12.8
Canada	8.6	14.3	Norway	1.4	10.8
Ceylon	1.6	19.8	Peru	−0.4	10.2
Chile	3.9	33.1	Philippines	5.3	40.9
Denmark	5.4	11.2	Portugal	4.5	11.6
Finland	7.2	22.8	South Africa	4.1	22.3
France	10.6	22.5	Spain	−1.5	17.7
Germany	16.6	28.3	Sweden	7.4	13.1
India	−1.0	13.7	Switzerland	6.3	20.5
Israel	7.4	37.0	United Kingdom	7.2	13.2
Italy	11.1	21.8	United States	12.1	12.1
Japan	17.8	31.3	Venezuela	4.2	16.8

Source: Calculated from common stock indices and exchange rates for each of the countries appearing in various issues of the *IMF*'s *International Financial Statistics*.

$P_{i(t)}$ = the dollar value of the ith country's common stock index at the end of year t.

$r_{i(t)}$ = the rate of return in year t.

The returns have been calculated in dollars and, therefore, the indices have been adjusted to reflect any changes in exchange rates during the period.[4] As a result, the rates of return used in this study are the relevant rates for investors in dollars or any other currency unit whose dollar rate of exchange remained constant during the period 1951–67.[5]

The mean rate of return for each country was calculated by taking the arithmetic average[6] of the annual returns:

$$R_i = \frac{1}{N}\sum_{t=1}^{N} r_i(\mathrm{t}) \tag{2}$$

A correlation matrix among the annual rates of return of the 28 countries making up our population was calculated and the variances and covariances were used in all subsequent computations. The variance for the ith country is defined as:

$$\sigma_i^2 = \frac{1}{N}\sum_{t=1}^{N} (r_i(\mathrm{t}) - R_i)^2 \tag{3}$$

The annual average dollar rate of return on common stocks for the period 1951–67 ranged from a high of 17.8 percent for Japan to a low of −1.5 percent for Spain. The intercountry differences in risk (standard deviations) were also very pronounced: four countries, Chile, Israel, Japan and the Philippines, had standard deviations greater than 30 percent; while seven countries had standard deviations below 13 percent.

As might have been anticipated, the rate of return for the United States (12

[4] Foreign currency values were converted into dollar values using the exchange rates which prevailed at the end of each year. Data on exchange rates were taken from various issues of *International Financial Statistics*.

[5] Thus, for example, the optimal investment proportions set out in this paper are relevant for Switzerland as well as for the United States, but not for the United Kingdom. Parallel studies of portfolio relationships using the relevant rates of return for other countries, e.g., United Kingdom, France, Israel and Japan, will be the subject of a separate paper.

[6] The rates of return have a downward bias owing to the neglect of dividends, however, the use of the arithmetic, rather than the geometric, mean imparts an offsetting upward bias to the historical rates of return.

percent) was relatively high while the risk level was relatively low (the standard deviation was also 12 percent); but, despite the relatively good performance of *U.S.* common stocks, American investors may still benefit from international diversification. It has been shown that as long as the correlation of returns among investment options is not perfect, a necessary, but not sufficient, condition for portfolio diversification exists (see Markowitz (1952, 1959) and Tobin). Thus the addition of even relatively low return foreign stocks might materially reduce the variance of the overall portfolio.[7]

In order to make an empirical test of the benefits to the investor from international diversification, we must first calculate the set of efficient portfolios; an efficient international portfolio being defined as a combination of investments in various countries which either maximizes the rate of return given the variance, or minimizes the variance given the rate of return. The locus of all such points comprises the efficiency curve, with each point on the curve representing a particular combination of investment proportions in various countries.[8]

The locus of efficient points (portfolios) was found by deriving the investment proportions[9] X_i which minimize the variance of the portfolio, for given expected rates of return. Formally the problem can be stated as follows. Find the vector X which minimizes the objective function C:

$$C = \underline{X}' \sum \underline{X}, \tag{4}$$

subject to the following restraints:

$$X_i \geqq 0 (i = 1, 2, \ldots, 28) \tag{5}$$

$$\underline{X}'\underline{R} = E \tag{6}$$

$$\underline{X}'\underline{1} = 100 \tag{7}$$

Where: X_i denotes the proportion of the portfolio invested in the ith country and Σ denotes the variance-covariance matrix of the rates of return on investment in countries $i, j (i, j=1, 2, \ldots, 28)$, and R_i denotes the average rate of return on investment in the ith country so that $\underline{X}'\underline{R}$ represents the portfolio rate of return for a given vector of investment proportions $\underline{X}$. The final constraint, $\underline{X}'\underline{1}=100$, ensures that the investment proportions of the international portfolio add to 100 percent.

The optimal vector $\underline{X}$ for various rates of return, was found by means of quadratic programming.[10] Minimum variances for alternative levels of portfolio return were determined by raising the given portfolio return by half a percent and repeating the computation procedure.

II

The efficiency frontier, i.e., the locus of efficient portfolios, for the case in which investments can be made in all of the 28 countries included in the study is given in Figure 1. The curve, labeled A, summarizes the efficient risk-return combinations which were attainable to an investor who had the opportunity to build an internationally diversified portfolio of common stocks. However, in order to focus attention on the proportion of investment in each of the various countries, a method must be found to reduce the efficiency locus to a single point (portfolio). Essentially this means that given a particular

[7] Moreover, three countries, Austria, Germany and Japan, had rates of return which exceeded that of the United States during the relevant period.

[8] A similar efficiency curve for a smaller group of countries over a shorter time period was constructed by Herbert Grubel.

[9] The computation procedure restricts the range of proportions to $0 \leqq X_i \leqq 1$, i.e., negative investment is not permitted. This is equivalent to assuming the absence of short sales.

[10] The optimal vector was found using the reduced gradient algorithm, *PHIMAQ*, developed by Electricité de France. The computations were made on the *CDC* computer at the Computation Center, Hebrew University.

set of initial conditions we must chose the optimum portfolio.

This can be accomplished by utilizing the market equilibrium model developed by John Lintner and William Sharpe. This model is illustrated in Figure 1.

The market opportunity line, r_1, rising from the 3 percent intercept on the Y-axis reflects the assumption that lending or borrowing at 3 percent can take place; similarly, the market opportunity line, r_2, reflects the case where borrowing or lending can take place at 6 percent.[11] The optimum unlevered portfolio for a particular interest rate is given by the point at which the appropriate market opportunity line[12] is tangent to the locus of efficient portfolios, e.g., points such as *a* and *b* in Figure 1. By constructing similar market opportunity lines, a one to one correspondence can be established between the interest rate confronting investors and an optimal point on the efficiency locus.[13]

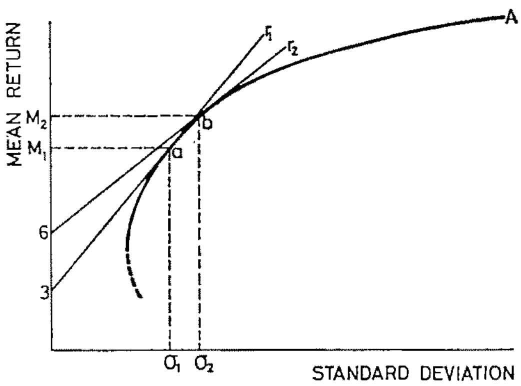


FIGURE 1

Table 2 sets out the investment proportions of four such optimal portfolios for 2, 3, 4, and 6 percent interest rates, respectively. Although twenty-eight countries are included in the study, only nine countries are included in at least one of the optimal portfolios in the relevant range, and of these, two countries (the United Kingdom and Denmark) can be ignored since only a negligible proportion of one of the portfolios is invested in these countries. Investments in the United States and Japan account for a majority (50 to 70 percent) of the optimal portfolios, but with the exception of Austria the portfolios virtually exclude the developed countries of Western Europe. Perhaps the most striking feature of the composition of the diversified international portfolios is the relatively high proportion of investments in developing or borderline income countries such as Venezuela, South Africa, New Zealand, Mexico, and Japan. Depending on the interest rate assumed, the proportion of such investment accounts for about 40 to 60 percent of the aggregate portfolio.

An explanation for these results is not difficult to find. Table 3 sets out the correlation matrix for the countries which were included in at least one optimal portfolio. Japan, which taken by itself is characterized by a high return but also by a

[11] The two market opportunity lines can also be used to illustrate a case of an imperfect capital market in which the borrowing rate differs from the lending rate.

[12] The slope of the market opportunity line measures the trade-off between expected return and risk for all investors, independent of their tastes, and, therefore, uniquely determines the optimal unlevered share portfolio for all investors faced by the relevant interest rate.

[13] Figure 1 sets out the market opportunity lines and tangency points for 3 percent and 6 percent interest rates. It is clear from the slope of the efficiency locus that the higher the interest rate (i.e., the higher the intercept of the market opportunity line on the Y-axis) the higher will be the point of tangency along the efficiency curve. We can conclude, therefore, that investors faced with a high interest rate will prefer an investment portfolio with both a higher risk and a higher average return than that chosen by an investor who is confronted by a relatively low interest rate.

This preference for riskier portfolios where higher interest rates prevail may not be intuitively obvious, but it readily can be explained. A higher interest rate is equivalent to establishing a lower price for a unit of risk (i.e., the slope of the 6 percent market opportunity line, r_2 is flatter (lower) than that of the 3 percent line, r_1 (see Figure 1). And as is true for any normal commodity whose price is lowered, more is consumed. In other words, given the higher interest rates the investor can add risk to his portfolio at more advantageous terms, and he, therefore, is prepared to absorb more variance to acquire additional increments of return.

TABLE 2—COMPOSITION OF OPTIMAL PORTFOLIOS FOR SELECTED INTEREST RATES
(in percent)

Country[a]	Interest Rates 2	3	4	6
Austria	3.43	6.99	9.03	12.06
Denmark	2.06	0.01	—	—
Japan	14.68	16.71	17.65	20.86
Mexico	4.03	4.32	4.53	—
New Zealand	13.16	6.27	2.59	—
South Africa	10.37	12.51	13.83	12.86
United Kingdom	0.18	—	—	—
United States	36.57	40.99	42.79	51.06
Venzuela	15.51	11.72	9.57	3.16
Total	100.00	100.00	100.00	100.00
Mean Portfolio Rate of Return	9.5	10.5	11.0	12.5
Portfolio Standard Deviation	5.72	6.41	6.82	8.39

[a] The remaining nineteen countries are not included in any of the optimal portfolios.

high level of risk, has a relatively large share in the portfolio owing to a high degree of negative covariance with the other members of the set. The return on Japanese investments is negatively correlated with five countries, while the correlation with the remaining three countries is close to zero. Thus, despite Japan's relatively high standard deviation, because of covariance, its inclusion tends to reduce the risk of the overall portfolio. This is especially true since the investment returns of Japan are negatively correlated with those of the United States, and the latter comprises a very large share of all the optimal portfolios. The United States, on the other hand, enters the optimal portfolios essentially on its own merits, the American investment returns having both a relatively high rate of return and only a moderate risk level.

III

The potential gains from international diversification can be seen more clearly by examining Figure 2 which again sets out the efficiency frontier for the 28 countries included in our population as well as the

TABLE 3—INTERCOUNTRY CORRELATION COEFFICIENTS FOR COUNTRIES INCLUDED IN OPTIMAL INVESTMENT PORTFOLIOS

	1	2	3	4	5	6	7	8	9
1. Austria	1.00	0.10	0.05	−0.14	0.29	−0.51	0.38	0.26	0.15
2. Denmark	—	1.00	0.04	0.17	0.53	−0.09	0.26	0.19	−0.17
3. Japan	—	—	1.00	−0.24	−0.20	−0.36	0.04	−0.26	−0.09
4. Mexico	—	—	—	1.00	0.22	0.10	−0.18	0.02	0.21
5. New Zealand	—	—	—	—	1.00	0.18	0.59	0.08	−0.32
6. South Africa	—	—	—	—	—	1.00	0.05	0.08	−0.08
7. United Kingdom	—	—	—	—	—	—	1.00	0.26	−0.32
8. United States	—	—	—	—	—	—	—	1.00	−0.17
9. Venezuela	—	—	—	—	—	—	—	—	1.00

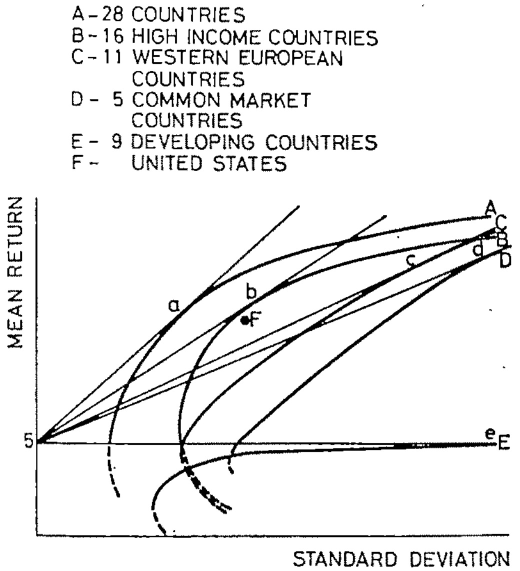


FIGURE 2

curves for several subsets[14] of the population. The risk-return combination for a portfolio invested in a cross-section of *U.S.* common stocks is represented by point *F* (rate of return and standard deviation both equal to 12.1 percent). Clearly, an investor will suffer a loss if he must eschew investment in the United States and restrict his portfolio to either developing countries, the common market, or Western Europe, all of which curves lie below point *F*.[15] Perhaps even more striking is the fact that by diversifying his portfolio to include investments in other high income countries as well as in the United States, the investor can only make a very marginal improvement in his portfolio, since point *F* lies very close to the efficiency locus *B*. Thus while diversification always pays, it does not pay very much in this instance.

It is only when the American investor diversifies his portfolio to include such countries as Japan and South Africa and the developing countries of South America and Asia that a significant improvement in his portfolio results; the efficiency locus for all the 28 countries included in the study (curve *A*) lies considerably to the left of point *F*.

TABLE 4—MEAN RATES OF RETURN AND STANDARD DEVIATIONS OF OPTIMAL PORTFOLIOS FOR A 5 PERCENT INTEREST RATE

(in percent)

	Mean Rate of Return	Standard Deviation
Developing Countries	5.0	26.5
Common Market	15.5	25.0
Western Europe	15.5	23.5
High Income Countries	13.0	12.5
All countries	12.0	8.0

The gains from diversification are quantified in Table 4 which sets out the mean rates of return and standard deviations of the optimal portfolios for each subset assuming a 5 percent interest rate.[16] The systematic nature of risk reduction through international diversification is reflected in the continuous reduction of the portfolio variance (at all levels of return) as the opportunity set is broadened. Thus, the best combination that can be created out of equities in the developing countries is a portfolio with a 5 percent return and a 26.5 percent standard deviation as compared with a return of 12 percent and standard deviation of 8 percent for the unconstrained optimum portfolio.

IV

Although the American investor should

[14] Three borderline countries; Israel, Japan, and South Africa, which are included in the total population, are not included in any of the subsets of Figure 2.

[15] This is not necessarily true according to the simple mean-variance criterion since some of the points on these curves represent efficient points relative to the United States. It is clear, however, that once borrowing and lending can take place, the investor can always reach a higher market opportunity line (and therefore a higher level of utility) by moving to point *F*, independent of the interest rate assumed. Cf. Lintner and Sharpe.

[16] The data of Table 4 correspond to the tangency points, *a*, *b*, *c*, *d*, and *e* of Figure 2.

TABLE 5—INTERCOUNTRY CORRELATION COEFFICIENTS FOR FIVE COMMON MARKET COUNTRIES AND THE UNITED STATES

	Belgium	France	Germany	Italy	Netherlands	U.S.
Belgium	1.00	0.66	0.60	0.31	0.65	0.83
France	—	1.00	0.46	0.60	0.49	0.34
Germany	—	—	1.00	0.71	0.76	0.43
Italy	—	—	—	1.00	0.59	0.09
Netherlands	—	—	—	—	1.00	0.53
United States	—	—	—	—	—	1.00

never restrict his portfolio to developing countries alone, (the efficiency curve E lies far below point F), the inclusion of these countries in the opportunity set materially improves his risk-return position. This has some interesting implications for the theory of international capital movements.[17] For example, the traditional approach to international investment which compares returns in developing countries with those of the developed economies of the United States and Europe, understates the benefits of such investments. For as we have seen, low yielding foreign investments in the developing countries may have a salutary effect on the overall portfolio variance of the investing country. Thus, when stabilizing portfolio effects are taken into account, a much stronger case can be made for investment in countries whose economies are not highly correlated with that of the investing country.

This is not the case for investments in the common market countries or Canada, since these countries are not included in the optimal portfolios. Table 5 helps clarify the reasons for this; the investment returns in the five common market countries included in Table 5 are all highly correlated so that little gain can be realized from combining them in a portfolio of their own.[18] This can also be seen in Figure 2 where the efficiency curve of the common market countries approximates a straight line. But even granted the high degree of positive correlation among the common market countries, this would not necessarily preclude the inclusion of at least one of the countries in the optimal portfolio. The last column of Table 5, however, shows that all of the common market countries have positive correlations with the United States as well.

It is also interesting to note the absence of Canada from the optimal portfolios. Again the main reason stems from that country's very high positive correlation (0.81) with the United States, combined with the fact the rate of return on Canadian investments was lower while the risk in Canada was greater than that for the United States. As a result, the United States dominates Canada and the latter was eliminated from the efficiency curve in the relevant range.

The composition of the optimal portfolios also raises some fundamental questions regarding the degree of imperfection in the international capital markets. In the absence of artificial barriers, we would expect an optimal portfolio (based on *ex*

[17] In a recent article, C. H. Lee has applied the portfolio principle to the analysis of capital movements between the United States and Canada. In the case of capital flows, the direction of the net capital movement depends not only upon the proportion of investment in each country's securities but upon the relative size of their captial markets as well.

[18] The high degree of correlation constitutes impressive evidence of a high degree of economic integration among the capital markets of these countries.

ante predictions)[19] to contain all countries' securities; if a country is not included, its share prices would fall (and rates of return rise) to levels where it would be included in the optimal portfolio. The *ex post* results, which show other combinations to be dominant, suggest that restrictions on international trade and/or capital flows have a significant effect on the pattern of security returns and permit inefficient markets to persist.

[19] The empirical findings are based on *ex post* data, and we do not purport to predict future returns. Moreover, virtually no theory is available to assess the degree to which the *ex post* results mirror *ex ante* predictions.

REFERENCES

H. G. Grubel, "Internationally Diversified Portfolios: Welfare Gains and Capital Flows," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1299–1314.

C. H. Lee, "A Stock Adjustment Analysis of Capital Movements: The United States—Canadian Case," *J. Polit. Econ.*, July/Aug. 1969, *77*, 512–23.

J. Lintner, "Security Prices, Risk and Maximal Gains from Diversification," *J. Finance*, Dec. 1965, *20*, 587–615.

H. M. Markowitz, "Portfolio Selection," *J. Finance*, Mar. 1952, *6*, 77–91.

———, *Portfolio Selection*, New York 1959.

W. F. Sharpe, "Capital Assets Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. Finance*, Sept. 1964, *19*, 425–42.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. Econ. Stud.*, Feb. 1952, *26*, 65–86.

Federal Reserve Bulletin, various issues.

International Monetary Fund, *International Financial Statistics*.

Soviet Postwar Economic Growth and Capital-Labor Substitution

MARTIN L. WEITZMAN*

The postwar record of the Soviet economy constitutes a largely unexplored case study for the measurement of aggregate production functions and their application to explaining economic growth. In the Soviet case high growth rates of capital and the absence of business cycles combine to create a relatively rich sample of full employment capital-labor and capital-output ratios. This results in greater statistical power than is typically available in the records of most Western economies for testing the appropriate *form* of an aggregate production function.

Because of weak theoretical underpinnings the aggregate production function at best yields an incomplete or even an ambiguous picture of an economy. The absence of appropriate aggregation theorems notwithstanding, the aggregate approach may be one of the few useful ways of piecing together an overall account of economic growth. It is in this spirit, rather than with misplaced notions of rigor, that aggregate production function identification and applications are undertaken in the present paper.

Granted that the somewhat nebulous concepts of aggregate output, capital, and labor have been fairly represented by the values cited in this paper and that functional forms have been correctly specified, I will argue that aside from a geometric time trend the postwar Soviet growth record is adequately accounted for by a constant elasticity of substitution (*CES*) production function with elasticity of substitution significantly less than one. Such a finding has certain implications for past and future Soviet growth, some of which are discussed. Results of a few alternative specifications are reported. Where they are relevant, some comparisons are made with the record of *U.S.* postwar economic growth.

I. *Basic Facts of Soviet Postwar Growth*

The primary series utilized in this study is displayed in Table 1. The estimates are for "industry" (mining, manufacturing, power) from 1950 to 1969 and rely directly on Soviet data. An industrial output index is obtained by aggregating official Soviet constant price gross value of output series for industrial sectors, using synthetic 1960 value-added weights. Official Soviet figures on "industrial productive basic funds" (July 1, 1955 prices) form the index of gross fixed industrial capital stock. Industrial man-hours worked forms the labor series.

Mainly to test the invariance of results with respect to somewhat different indexes, a series which is primarily of Western origin was also examined. Table 2 presents estimates of output, fixed capital services (gross), and man-hours worked for the nonagricultural nonservice economy (industry, construction, transportation and

* The author is associate professor of economics at Yale University. He would like to thank James Noren for the detailed and very useful comments which he provided. He is also indebted to Raymond Powell for his advice on many aspects of Soviet statistics. Neither of them should be blamed for the numbers used or the conclusions drawn from their use. The research described in this paper was carried out under grants from the National Science Foundation and the Ford Foundation.

TABLE 1—INDEXES OF OUTPUT, CAPITAL, AND LABOR FOR SOVIET INDUSTRY, 1950–69

(Mining, Manufacturing, Power—Direct Soviet Sources, 1960=100)

Year	Y	g_Y[b]	K	g_K[b]	L	g_L[b]
1950	33.15		33.77		79.92	
1951	38.20	15.2	37.59	11.3	84.52	5.8
1952	42.56	11.4	42.35	12.6	87.60	3.7
1953	47.35	11.3	46.96	10.9	91.41	4.3
1954	53.46	12.9	52.18	11.1	95.83	4.8
1955	60.34	12.9	58.64	12.4	97.61	1.9
1956	66.33	9.9	65.64	11.9	96.32	−1.3
1957	73.01	10.1	72.50	10.5	96.60	.3
1958	81.29	11.3	80.45	11.0	99.14	2.6
1959	90.56	11.4	89.67	11.5	100.66	1.5
1960	100.00	10.4	100.00	11.5	100.00	− .7
1961	109.56	9.6	111.59	11.6	99.54	− .5
1962	120.27	9.8	123.93	11.1	102.81	3.3
1963	131.21	9.1	138.10	11.4	106.44	3.5
1964	141.78	8.1	154.31	11.7	110.91	4.2
1965	153.41	8.2	170.42	10.4	115.65	4.3
1966	167.09	8.9	186.13	9.2	118.66	2.6
1967	183.65	9.9	201.73	8.4	122.55	3.3
1968	199.91	8.9	217.68	7.9	126.62	3.3
1969[a]	215.2	7.6	235.1	8.0	130.6	3.1

[a] Preliminary.
[b] Annual percentage rate of growth.

TABLE 2—INDEXES OF OUTPUT, CAPITAL, AND LABOR FOR THE SOVIET UNION, 1950–66

(Aggregated Industry, Construction, Transportation and Communications, and Distribution—Western Estimates, 1960=100)

Year	Y	g_Y[a]	K	g_K[a]	L	g_L[a]
1950	38.84		37.87		80.79	
1951	43.66	12.4	41.35	9.2	84.51	4.6
1952	47.47	8.7	45.13	9.2	87.09	3.1
1953	51.99	9.5	49.39	9.4	89.71	3.0
1954	57.80	11.2	54.30	9.9	93.87	4.6
1955	64.36	11.3	59.93	10.4	95.47	1.7
1956	70.26	9.2	66.55	11.0	93.65	−1.9
1957	77.60	10.4	73.93	11.1	94.50	.9
1958	84.71	9.7	81.86	10.7	98.10	3.8
1959	92.38	9.1	90.51	10.6	100.88	2.8
1960	100.00	8.2	100.00	10.5	100.00	− .9
1961	107.54	7.5	110.49	10.5	98.73	−1.3
1962	115.51	7.4	121.96	10.4	100.96	2.3
1963	122.76	6.3	134.22	10.1	103.87	2.9
1964	131.12	6.8	147.33	9.8	108.01	4.0
1965	141.96	8.3	161.27	9.5	112.38	4.0
1966	151.97	7.1	175.86	9.0	114.90	2.2

[a] Annual percentage rate of growth.

communications, and distribution) over the years 1950–66. Because the highly aggregated investment data of R. H. Moorsteen and R. P. Powell (1966, 1968) underlie the capital stock series of Table 2, I was forced to include all activities of the nonagricultural economy consistent with the principle of excluding service outputs which are measured directly in terms of factor inputs (like education or housing). The output series is aggregated over indexes of Western origin using synthetic 1956 value-added weights. Capital services are built up out of constant-price investment data.

A more detailed description of the data sources is reserved for the Appendix.[1] Having experimented with a variety of output indices (with various price bases, diverse interest rates, different imputations for value-added weights, etc.) and alternative measures of capital, I can report that results are not much changed. Limitations of space preclude a more detailed report of the alternatives, but there is not much difference.[2] The two series presented here were selected for their contrasting coverage and somewhat different origins. The first one is considered to be of primary interest because the scope is limited to industry, coverage is believed more accurate, and it extends over more years. Where not otherwise noted, specific regression results will refer to data from the first series, although in all cases I have checked that similar conclusions hold for the second set of data also.

Both series show similar trends. The rate of growth of output apparently has declined somewhat over time. Capital

[1] The Appendix is limited to explaining the origins of the primary series of Table 1. Derivation of the Table 2 indexes is discussed in the appendix of my discussion paper of 1968.

[2] Most of the difference shows up in the form of a slightly altered estimate of the rate of growth of technical change. Estimates of the other parameters remain about the same.

grows in a steady manner, also decelerating toward the end of the period. Man-hours employed grows erratically. The low growth rates of labor from 1956 to 1960 are mostly due to the reduction of the length of the work week. The growth of the labor force is curtailed in the early 1960's as a result of sharply lower wartime births. Capital-output ratios remain roughly constant or decline slightly up to the mid 1950's but rise more or less steadily thereafter. The trend of an increasing capital-output ratio would be more sharply established if one could expunge the effects of technical change from output.[3]

II. *Explanations of Soviet Growth*

The observation that a retardation in output growth has taken place without a slowdown of the same magnitude in the growth of inputs has been the starting point for several discussions of the residual's somewhat paradoxical behavior.[4] When output is divided by a (usually geometrically) weighted index of inputs, it is discovered that the growth of total factor productivity, while volatile from year to year, has declined over the long run. Of course the magnitude of productivity growth is influenced by the choice of factor weights, but almost no matter what figures are used, growth of the residual seems to decline over time. By process of elimination output deceleration is implicitly attributed to the stalled growth of efficiency, technical change, or whatever else is believed to stand behind the residual.[5]

It is important to bear in mind that this approach relies on what in effect are some hidden assumptions about the aggregate production function. Basically it is presumed that the expression

$$Y(t) = A(t)F[K(t), L(t)] \tag{1}$$

is a serviceable approximation relating aggregate output Y to aggregate capital K and labor L at time t. In order to get at the unknown residual $A(t)$, the form of F is explicitly postulated, usually to be a Cobb-Douglas production function with prescribed weights. In this paper an alternative specification is emphasized.

III. *Role of the Elasticity of Substitution in Explaining Soviet Growth*

A widely accepted measure of the ease with which K and L can be substituted for one another is the elasticity of substitution, σ, which is defined as follows when the production function is homogeneous of the first degree:

[3] Detrended capital-output and labor-output ratios for industry are calculated in the first two columns of Table 3 with $\hat{\lambda}$, the growth of technology, equal to approximately 2 percent per year.

[4] See, e.g., Bela Balassa, Norman Kaplan, Moorsteen and Powell (1966, 1968), and J. H. Noren.

[5] As an explanation for the decreased growth of the residual, Moorsteen and Powell stress the notion that after 1953, productivity at first rose very rapidly while the economy broke away from the "rigidities and irrationalities" of Stalinism. However, ". . . by the late 1950's the more manageable inefficiencies of the economy had been largely remedied, and resistance to further movement toward the frontier, at least within existing institutional arrangements, had become severe" (Moorsteen and Powell 1968, p. 9). Kaplan (p. 302) emphasizes possible post-1958 declines in organizational efficiency due to the increased complexity of resource allocation. These arguments may well contain important elements of truth. However such strongly declining growth rates of $A(t)$ as result from specifying $F[K, L]$ as Cobb-Douglas with assigned factor weights strike me as unlikely. First of all, technological progress is at least partially a result of research and development, which is best thought of as a round-about method of production typically favored as capital becomes more plentiful and real rates of return decline. Also, the Soviet literature has been paying steadily increasing attention to questions of efficiency since the mid-1950's. Whether this is reflected in practice is difficult to say, but the ideas of Liberman, Kantorovich and others would appear to have influenced economic policy to a certain degree, especially lately. These considerations would appear to justify an *increase* in the growth of technical change over the period of the 1950's and '60's. Unfortunately, the sad truth must be that we really have no good independent information on this question.

$$\sigma \equiv \frac{\frac{\partial F}{\partial L}\frac{\partial F}{\partial K}}{F\frac{\partial^2 F}{\partial K \partial L}}$$

A 1 percent change in the ratio of prices or in the marginal rate of substitution between capital and labor is associated with a σ percent change of factor ratios in the opposite direction. As is well known, a production function additive in the two factors has an infinite elasticity of substitution, the Cobb-Douglas has unit elasticity, and a fixed coefficients production function has zero elasticity of substitution.

Because there is a basic identification problem in distinguishing between $A(t)$ and $F[K, L]$, a unit elasticity of substitution production function and $A(t)$ declining in growth over time is not the only way of explaining the Soviet growth record.[6] If the elasticity of substitution for the function $F[K, L]$ were of the appropriate magnitude (it would have to be less than one), a slowdown in the growth of output could be accomodated with both exponential growth of A and with the capital and labor series presented.

This is most easily seen by differentiating (1) and dividing by Y to obtain

$$g_Y = g_A + \eta_K g_K + \eta_L g_L \tag{2}$$

The rates of growth of Y, A, K, L are denoted, respectively, by g_Y, g_A, g_K, g_L. Also

$$\eta_K \equiv \frac{K\frac{\partial Y}{\partial K}}{Y} = \frac{K\frac{\partial F}{\partial K}}{F}$$

and

$$\eta_L \equiv \frac{L\frac{\partial Y}{\partial L}}{Y} = \frac{L\frac{\partial F}{\partial L}}{F}$$

are the imputed competitive shares of capital and labor. Note that if $F[K, L]$ exhibits constant returns to scale, as we shall assume, then $\eta_K + \eta_L = 1$.

So long as the elasticity of substitution were less than unity, the condition $g_K > g_L$ would imply that η_L would increase over time, and η_K would decline.[7] It follows that g_Y would tend to decline over time due to the increasingly heavy weight of the slower g_L. This is in contrast with the Cobb-Douglas function where the factor share weights remain constant. Such an effect might be quite pronounced in the Soviet case because g_K is very high relative to g_L.

For the purposes of this paper, it is useful to think of the elasticity of substitution as a measure of the rate at which diminishing returns sets in as one factor is increased *relative* to the other. A less than unit elasticity of substitution implies eventual difficulty in increasing output by primarily incrementing one factor, because diminishing returns set in strongly and rapidly. Such a situation would have special relevance for the Soviet case because capital has grown so fast relative to labor.

[6] For a more explicit discussion of this point, see Marc Nerlove who also presents some interesting survey material on empirical measurement of the *CES* function.

[7] Proof: Let $k \equiv K/L$, and $f(k) \equiv F[K/L, 1]$. Then

$$\sigma = \frac{-f'(f - kf')}{kff''}, \quad \text{and} \quad \eta_L = \frac{f - kf'}{f}$$

(assuming constant returns to scale). Differentiating η_L,

$$\frac{d\eta_L}{dk} = \frac{-kf''}{f} - \frac{(f - kf')f'}{f^2} = \frac{-kf''}{f}\left(1 + \frac{f'(f - kf')}{kff''}\right) \gtreqless 0$$

according as $\sigma \lesseqgtr 1$. Naturally η_K $(= 1 - \eta_L)$ moves in a direction opposite to that of η_L.

IV. *Fitting a CES Production Function*

Because it is not possible to identify both $A(t)$ and $F[K, L]$, one cannot proceed further without assuming explicit functional forms. For reasons that will become clear, the constant elasticity of substitution (*CES*) production function was thought to be a superior specification for the production side.[8] This production function can be written in the form

$$F[K, L] = \gamma[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{-1/\rho} \tag{3}$$

with $\sigma = 1/(1+\rho)$ the (constant) elasticity of substitution.

On the side of the residual, I can hope that Hicks neutral disembodied exponential technical change

$$A(t) = e^{\lambda t} \tag{4}$$

is not too bad a description of reality.[9]

It goes without saying that I have no good a priori reasons for believing in this specification over and above some reasonable alternatives. A few other specifications were also tried, and some of these will be reported. But in the final analysis the relatively simple form

$$Y(t) = \gamma e^{\lambda t}[\delta K(t)^{-\rho} + (1 - \delta)L(t)^{-\rho}]^{-1/\rho} \qquad \rho = \frac{1 - \sigma}{\sigma} \tag{5}$$

was felt to explain the data as well as any other expression tested, and better than most.

Because (5) is non-linear in the parameters, identification is often accomplished by utilizing factor price data to estimate a linear form. For example, under perfect competition the relevant parameters can be estimated by regressing the *log* of labor productivity on the *log* of wages and on time. Personally, an aggregate production function has always impressed me as being a tenuous enough concept even without depending for its measurement on first derivatives and a marginal productivity theory of distribution. At any rate, this is out of the question as a meaningful alternative for the present model. Instead, equation (5) was estimated directly by using a non-linear regression program which chose parameter values to minimize the sum of squared logarithmic residuals,[10] defined as

$$\Phi = \sum_t \left\{ \log Y(t) - \left[\log \gamma + \lambda t + \frac{\sigma}{\sigma - 1} \log\left(\delta K(t)^{-\rho} + (1 - \delta)L(t)^{-\rho}\right) \right] \right\}^2, \qquad \rho \equiv \frac{1 - \sigma}{\sigma}$$

A logarithmic form was chosen because it seems sensible to assume a multiplicative error term in (5).

Results for the data of Table 1 (industry—direct Soviet sources, 1950–69) are[11]

[8] See Kenneth Arrow, et al.

[9] From here on I use the terms "residual," "total factor productivity," "technical change," and "technological progress" interchangeably. Residual is really the best term because it most aptly conveys the notion that what is being identified is *any* contribution to growth other than capital and labor as conventionally measured and combined subject to constant returns to scale.

[10] The I.B.M. Share program 309401 (revised 8/15/66) entitled "Least Squares Estimates of Non-Linear Parameters" was used. It is based on an algorithm devised by D. W. Marquardt. While this algorithm can at best obtain only a *local* minimum, converging to the same parameter values with different initial estimates makes me confident that the estimates given here in fact globally minimize the error sum of squared residuals. If the multiplicative errors associated with (5) are independently identically distributed log normal random variables with mean one, the parameter values chosen to minimize the sum of squared logarithmic residuals will also be maximum likelihood estimates.

[11] Numbers in parentheses denote the standard errors of the coefficients to a linearized regression about the optimal parameter values. Let the non-linear model be expressed in the general form $Z = g(X, \beta) + \epsilon$ with ϵ assumed normally distributed, Z dependent and

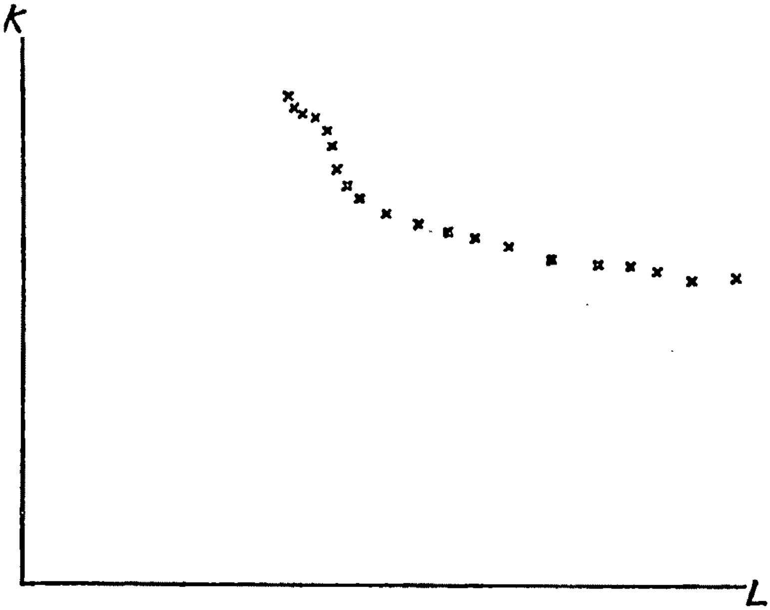


FIGURE 1. PRODUCTION ISOQUANT FOR SOVIET INDUSTRY, 1950–1969.

$$\hat{\sigma} = .403 \quad \hat{\lambda} = .0205$$
$$(.030) \quad (.0053)$$

$$\hat{\delta} = .639 \quad \hat{\gamma} = .804 \quad \bar{R}^2 = .9995$$
$$(.070) \quad (.044)$$

The outcome of this regression[12] can be interpreted to mean that given the specification (5), an elasticity of substitution significantly less than one has an important

X independent observations and β parameters. Let $\hat{\beta}$ be the minimizer of $\|Z-g(X, \beta)\|^2$. Standard errors reported are those of the least squares estimator of β in the linear regression model

$$Z = g(X, \hat{\beta}) + B(\beta - \hat{\beta}) + \omega,$$

ω normally distributed and $B = \partial g/\partial\beta|_{x,\hat{\beta}}$. The least squares estimate of β is $\hat{\beta}$. Standard errors so obtained, i.e., the diagonal terms of the matrix $[B^TB]^{-1}$, are valid only asymptotically as the sample size goes to infinity, or as the function g is close to being an inner product in X and β. I have looked into the accuracy of this approximation for parameters σ and λ by moving out two standard errors in both directions and ascertaining the increase in mean square error after subminimizing the error sum of squares over all possible values of the free parameters. The increase in mean square error was close to that predicted by the linearized model. Even the properties of this more refined test are large sample as far as using an F-statistic is concerned. With the present state of non-linear estimation theory there does not seem to be a more feasible alternative to assuming with tongue in cheek that large sample properties can be used in this case for hypothesis testing. As usual, $\bar{R}^2$ denotes the percentage variance explained by the (non-linear) regression, corrected for degrees of freedom. The Durbin-Watson statistic for this regression was $d = .81$. The low value of d was at first considered disturbing and the regression was rerun in weighted first differences to investigate a possible first-order autoregressive error. The coefficient point estimates and linearized standard errors hardly changed no matter what weights were chosen. The d statistic did not appreciably improve either. My tentative conclusion is that the low value of d coupled with a very high value of $\bar{R}^2$ indicates more the difficulty of parametric curve fitting than the presence of stochastic autocorrelation per se. It is a frequent tale in the folklore of empirical curve fitting that a very tight fit is often associated with a poor Durbin-Watson statistic. In this particular case the low value of d is due primaily to unexplained poor economic performance in the immediate pre-reform years 1964–66 rather than to the strong presence of autocorrelation throughout the postwar period. In any event, the reported linearized standard errors should be interpreted with even more caution than they otherwise might be.

[12] Results for the data of Table 2 (industry, construction, transportation and communications, distribution—Western estimates, 1950–66) were similar. They are:

role to play in explaining the postwar pattern of Soviet growth. To see this graphically, a scatter diagram of the implied industry production isoquant after technical change has been detrended from output is presented in Figure 1. The data are from the first two columns of Table 3. For each year (1950=0), values of

$$\frac{K(t)}{Y(t)e^{-\hat{\lambda}t}} \quad \text{and} \quad \frac{L(t)}{Y(t)e^{-\hat{\lambda}t}}$$

are plotted, with $\hat{\lambda}$ the least squares estimate of λ. The scatter isoquant of Figure 1 reveals the presence of significant diminishing returns.

TABLE 3—DETRENDED CAPITAL-OUTPUT AND LABOR-OUTPUT RATIOS AND THE IMPUTED SHARES OF CAPITAL AND LABOR FOR SOVIET INDUSTRY

Year	$\frac{K(t)}{Y(t)\, e^{-\hat{\lambda}t}}$	$\frac{L(t)}{Y(t)\, e^{-\hat{\lambda}t}}$	$\eta_K(t)$[a]	$\eta_L(t)$[a]
1950	1.018	2.411	86	14
1951	1.005	2.259	88	15
1952	1.037	2.144	84	16
1953	1.055	2.053	82	17
1954	1.059	1.946	81	19
1955	1.077	1.792	79	21
1956	1.119	1.642	75	24
1957	1.146	1.527	72	27
1958	1.166	1.437	70	29
1959	1.191	1.337	68	32
1960	1.227	1.227	65	37
1961	1.276	1.138	62	41
1962	1.318	1.093	59	44
1963	1.374	1.059	55	46
1964	1.450	1.042	51	47
1965	1.511	1.025	48	48
1966	1.546	.986	46	51
1967	1.556	.945	46	54
1968	1.575	.916	45	57
1969	1.612	.896	44	59

[a] Percent.

n. 12 ctd.

$$\hat{\sigma} = .274 \quad \hat{\lambda} = .0134 \quad \hat{\delta} = .587 \quad \hat{\gamma} = .875$$
$$(.018) \quad (.0029) \quad (.041) \quad (.026)$$
$$\bar{R}^2 = .9998 \quad d = 1.78$$

Some of the difference in estimates is due to unequal time coverage—cf Figure 1 without the last three years.

It can be shown that for the *CES* type production function (5),

$$\eta_K(t) \equiv \frac{K\dfrac{\partial Y}{\partial K}}{Y} = \delta\gamma^{-\rho}e^{-\lambda\rho t}\left(\frac{Y(t)}{K(t)}\right)^{\rho},$$

and

$$\eta_L(t) \equiv \frac{L\dfrac{\partial Y}{\partial L}}{Y} = (1-\delta)\gamma^{-\rho}e^{-\lambda\rho t}\left(\frac{Y(t)}{L(t)}\right)^{\rho},$$

where $\rho \equiv (1-\sigma)/\sigma$. The implied time-series of imputed factor shares for industry is shown in the last two columns of Table 3. Note that the sum $\eta_K(t)+\eta_L(t)$ fails to add up to exactly one because of the error term in the specification (5). In 1950 a 10 percent increase in capital would have increased output by almost 9 percent but by 1969 the output response has dropped to nearly 4 percent. Such results help to quantify a significant transformation of the Soviet economy from a near labor surplus situation in the early 1950's to a position now where labor scarcity is an important fact of economic life.

V. *Some Alternative Specifications*

This section tests the appropriateness of the *CES* production function with disembodied exponential Hicks neutral technical change against some alternative specifications. In all cases we use as the test statistic for a more restricted hypothesis the percentage increase in the error sum of squares (corrected for degrees of freedom) after reminimization under the more restricted hypothesis. This statistic will be asymptotically distributed as an F-statistic when the sample size is sufficiently large. Somewhat arbitrarily, we use

the 95 percent confidence level as our cutoff point.[13]

Forcing $\sigma=1$ in (5) yields the more restricted hypothesis of a Cobb-Douglas type production relation

$$Y = \gamma e^{\lambda t} K^{\delta} L^{1-\delta} \tag{6}$$

The error sum of squares (*ESS*) increases by 374 percent, easily rejecting the specification (6) at the 95 percent confidence level. This result was anticipated by our remarks about the strong curvature of the scatter isoquant in Figure 1.

In an effort to test for a systematic decrease in the rate of growth of the residual, the model

$$Y = \gamma e^{\lambda t + lt^2}[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-1/\rho} \tag{7}$$

was examined. The rate of growth of the residual, $\dot{A}/A$, is $\lambda + 2lt$ which would decrease over time provided l were negative. The model (7) is one particular parametrization for testing the more usual explanation of the slowdown of the Soviet economy in terms of a decreasing growth rate of the residual.

The more restricted hypothesis $l=0$, which reduces (7) to (5), increases the *ESS* by 22 percent. At the 95 percent confidence level, we cannot reject the hypothesis that (5) is the true functional form. On the other hand, the more restricted hypothesis $\sigma=1$ increases the *ESS* by 89 percent, rejecting by a large margin at the 95 percent confidence level the hypothesis that

$$Y = \gamma e^{\lambda t + lt^2} K^{\delta} L^{1-\delta}$$

is the true functional form. Thus, we are inclined to say that (5) offers a better explanation of the data than does the more usual explanation in terms of a Cobb-Douglas production function with decreasing growth of the residual (at least with a residual of the form $A(t) = e^{\lambda t + lt^2}$.

Also tested in an effort to probe the appropriateness of specification (5) were the following three models: (i) two separate rates of factor augmenting disembodied technical change—labor augmenting at rate λ_1 and capital augmenting at rate λ_2; (ii) uniform returns to scale in which the right-hand side of (5) is raised to the power μ; (iii) variable elasticity of substitution of the form $\sigma = \sigma_0 + \sigma_1 t$. In all three cases the more restricted hypothesis that (5) is the true functional form could not be rejected at the 95 percent confidence level.[14]

VI. *A Few Comparisons with the American Economy*

In order to compare the results obtained here with those for the *U.S.* economy, a direct estimate of a production function of the form (5) for American manufacturing data was made.[15] (I have also included a multiplicative term of the form $e^{\theta U}$, U = unemployment, θ = a parameter to be estimated, for some of the regressions). Although the data covered the years 1919 to 1967, various subperiods were also tried (especially 1950–67), and runs were made omitting recession years and/or war years.

[13] Let n be the number of observations and k the number of coefficients to be estimated under the hypothesized functional form H. Let H' be the more restricted hypothesis that the i^{th} coefficient β^i equals $\bar{\beta}_i$. Let *ESS* be the error sum of squares. The statistic

$$\psi \equiv (n-k)\frac{ESS_{H'} - ESS_H}{ESS_H}$$

will be asymptotically distributed as an F-statistic with $(1, n\text{-}k)$ degrees of freedom. We fail to reject H' if ψ is below the upper 5 percent point of $F_{1,\,n-k}$. The proposed F-test is equivalent to a likelihood ratio test.

[14] The existence of embodied capital augmenting technical change was also examined. Results were disappointing because the reduction in the *ESS* was insignificant. For details, see Weitzman.

[15] The coverage of the American and Soviet data are not entirely comparable. Soviet industry includes power, mining, fishing, and logging, all of which are excluded from *U.S.* manufacturing. The capital series are from Robert Gordon, updated by Commerce figures. Output is from Gordon, based on Kendrick's and *O.B.E.* data, and updated by the latter. The labor series is man-hours worked in manufacturing, *B.L.S.*, based on establishment data. Unemployment rate is from *O.B.E.* data.

The general picture which emerged did not seem to be terribly dependent upon which permutation of *ad hoc* adjustments was used.

Even the best fits were generally worse than in the Soviet case (standard errors about three times as large.) This is undoubtedly largely due to the much greater stochastic element in *U.S.* production because business cycles cause capacity utilization to vary a great deal. It may also be due, at least in part, to the relatively more important role of technical change. This role, it might be reasonable to assume, is probably one of the most poorly specified parts of the production function model. One might suspect that a model reflecting a situation with greater growth of technical change relative to output growth would inherently possess greater output variance because much of the "error" could be due to the undetected swings of technical change, now much wider, in an unobserved variable merely assumed to grow smoothly over time.

Estimates of λ were uniformly close to each other in every regression (about 2.3 percent per year with standard errors of only about .2 percent). However, the elasticity of substitution has such a large asymptotic standard error that it looked as if it could easily have been anywhere between 0 and 1.5. This was partially verified by directly fixing a few values of σ and subminimizing the error sum of squares. Thus, for example, $\sigma=.3$ gave as good a fit as $\sigma=1$. It was also observed in a scatter diagram analogous to Figure 1 that the hoped for "isoquant" really looked more like the interior of an ellipse. There is simply too much autocorrelation (and too high a standard error) in the *U.S.* data to be able to pull out even a reasonably precise estimate of the elasticity of substitution, at least by the methods employed here. If capital and labor grow at comparable rates, the growth of output is not much affected by the elasticity of substitution, which can hardly be expected to be measured with precision in such situations. The Soviet case, with its greater variety of input configurations and less erratic growth record affords a much better experiment for measuring a second-order parameter of curvature like the elasticity of substitution.

It is instructive to ask why the Soviet economy has exhibited a slowdown while nothing comparable has occurred in the *U.S.* economy. This is undoubtedly a question deserving of fuller study, but in the spirit of the analysis carried out here several answers suggest themselves. Although no statistical evidence could be found one way or another, the *U.S.* elasticity of substitution may well be greater than the Soviet, so there is little to explain along the lines of the argument that diminishing returns should have caused a slowdown. Or, technical change in the United States may really be labor augmenting. However, even if these differences and several others did not exist, there would still remain important contrasts between the two economies. In Equation (1), λ is such a larger percentage of g_Y in the American case that the overall effect of changes in η_K (or η_L) on output are reduced. Furthermore the American capital-labor ratio grows so much slower relative to the Soviet that changes in η_K and η_L would tend to be much more gradual.

As a matter of fact, in situations where capital does not grow appreciably faster than labor, little is to be gained by going over from the Cobb-Douglas to the *CES* model because the drag due to even a low elasticity of substitution would be minimal.[16] A major finding of the present paper is that this is not the appropriate case for the Soviet postwar economy.

[16] See Richard Nelson who carried out a Taylor series expansion of a *CES* function in order to demonstrate this conclusion.

VII. *Conclusion*

Data on Soviet industrial output, capital, and labor have a distinct air of diminishing returns about them. This is clearly visible in the detrended isoquant scatter diagram of Figure 1. Given the structure which was imposed by equation (5), the regression results point to an identifiable elasticity of substitution which is significantly less than one. The rate of growth of technical change, estimated at about 2 percent per year, is respectable if not spectacular by world standards.

What kind of story is told by these findings? Accounting for somewhere around 15 to 25 percent of average output increases, technical change is not nearly so significant a determinant of economic growth as in some other economies. The Soviet postwar record looks very much like a classical model of economic growth with high rates of capital accumulation serving as the prime mover.

A low elasticity of substitution makes labor look in the early 1950's as if it were practically surplus, output going up almost proportionally with capital. By the late 1960's this is no longer the case, as is obvious from the tabulations of η_K and η_L in Table 3. By this time a low elasticity of substitution seems to imply that capital accumulation has outstripped labor growth by a wide enough margin that the drag due to diminishing returns is significantly cutting into output growth. In the mature Soviet economy, labor scarcity appears to be a reality.[17] We again stress that the present emphasis on diminishing returns is very different from the somewhat more usual factor productivity approach. In that story the sharply diminished growth of factor productivity usually emerges as the main reason for the Soviet slowdown.

Having come this far with a *CES* model, there is a great temptation to speculate on the implications for future Soviet growth. Unfortunately, there has to be a wide margin of error associated with even short-run projections of this type. But *if* Soviet industry is accurately described by the aggregate production function (5), *if* the parameter values have been accurately identified, and *if* they do not change very much in the near future, then it would appear that a strategy of strong capital accumulation must be considerably less successful for the present relatively mature Soviet economy than for its labor surplus predecessor.

Although a continuation on the same scale of a strategy of capital deepening can still yield growth rates which are high by Western standards, the days of relying almost exclusively on capital formation for producing 10 to 15 percent annual increases in industrial output would appear to be over.[18] Instead of capital, labor and technical change will have to be increasingly relied upon as alternative sources of future economic growth.

Labor is more important these days because the great past accumulations of capital have combined with a sufficiently low elasticity of substitution to increase its marginal product dramatically. But can the growth of man-hours be stepped up? Probably not. Demographers estimate that the growth of the working age population will not increase in the near future.[19] Nor can industrial laborers be so easily drawn out of agriculture as they might have been in the past. It seems safe to say that the growth of the industrial labor force cannot

[17] A. Nove (p. 167) believes it a fair current generalization to say that "the period of labor abundance is drawing to a close." He goes on to cite Strumilin, who attributes the new interest of Soviet economists in obsolescence (a concept whose very applicability to the U.S.S.R. was denied before 1955) to the change in the relative scarcities of capital and labor.

[18] In any case, it does not appear likely that the previous scale of capital accumulation will be kept up, since the rate of growth of capital has slackened lately.

[19] See, e.g., M. Feshbach, pp. 713–14.

in the near future rise significantly above what it has averaged in the past five years.

This rests the spotlight, finally, on technical change. If the Soviets want to continue high rates of growth into the future, they will have to encourage more actively the residual element.[20] In terms of equation (2), the most appealing way of raising g_Y is now to increase $g_A(=\lambda)$ because g_L is more or less fixed and the low (relative to the past) current value of η_K now dampens the effects of large g_K. With increased λ, the Soviet growth scenario would resemble a little more closely the contemporary expansion of Japan and a few Western countries. The distinctive feature of high rates of growth of capital and output would likely remain but technical change would be doing more of the pushing.

This line of reasoning opens up some fascinating, if highly speculative issues. For example, there has always been a suspicion that Soviet emphasis on yearly, quarterly, and monthly plan fulfillment leads to a fear of uncertainty which has discouraged innovation on the local level. Does this mean that a greater degree of local autonomy on issues related to innovation and risk taking would help increase growth of the residual? More generally, what economic changes, if any, might be needed to coax out increased rates of technical progress? With their demonstrated commitment to rapid economic growth, the Soviet leaders may well continue their recent policy of hammering out those pragmatic organizational compromises considered necessary to secure future growth.[21] If so, the question of what administrative structure would be most conducive to stepped up technical change looms as an important one for Soviet economic society.

[20] It is at this point that our ignorance of what constitutes the residual becomes really annoying. What is it that should be pushed—increasing returns, labor skills, new inventions, optimal use of resources, better organization, or what? Naturally, I evade the issue. A few courageous researchers have tried to make a more specific breakdown of the residual (see, e.g., E. R. Brubaker). Their approach is suggestive, but an overpowering amount of *ad hoc* calculation is sometimes involved in cooking up the relevant figures.

[21] I am thinking mostly of the "New Economic System." The main emphasis of this reform has been on increased efficiency in combining resources and on the quality of production. This is likely to result in change of the static once-and-for-all variety which does not really get at the issue of increasing the residual over time. Effects of this reform can be read (or is it our imagination?) in some of the "wiggles" of the Figure 1 isoquant.

Appendix

Data Sources and Compilation

The purpose of this Appendix is to explain in some detail compilation of the labor, capital, and output series for Soviet industry which appear in Table 1. Limitations of space preclude a similar treatment of the data of Table 2, but the interested reader will find an account in the appendix of Weitzman.

Almost all of the raw data is obtained from Soviet statistical handbooks. As is to be expected, figures appearing in one year are sometimes later revised, typically by an insignificant percentage. The general principle followed here is that data from the more recent publication takes precedence in cases of duplication. When a date is associated with the annual *Narkhoz* (*Narodnoe Khoziastvo SSSR*) or *Tsifrakh* (*SSSR v Tsifrakh*) volumes, it always specifies the date to which the yearbook pertains, *not* the date of publication (which is almost always the following year).

The starting date for the present study was not made earlier than 1950 to be sure of not getting involved with the erratic postwar recovery. By 1950 industry was producing at well above prewar output levels.

Labor

The industrial man-hours per year series of Table 4 forms the labor input index of Table 1. For each year it is calculated as the product of the number of industrial

workers and personnel with the average number of days worked per year with the average length of the working day.

The number of industrial workers and personnel is based on a new series which appeared with the 1968 *Narkhoz* (p. 548) and differed slightly in coverage from the series presented in previous yearbooks. Only the years 1960 and 1965–68 were covered in the '68 *Narkhoz*. Fortunately Murray Feshbach was kind enough to provide me with the complete new series from 1958 to 1968 inclusive (this series will be cited in a forthcoming Joint Economic Committee study on the U.S.S.R.). In 1958 the difference between the "new" and the "old" series is .86 of one percent, and it is undoubtedly less for earlier years. I have taken the difference between the new and the old series in 1958 (181,000) and reduced it by a multiplicative factor of 2/3 for each year preceding 1958, adding the result to the old series of *Trud v SSSR* (p. 81). The 2/3 factor is the ratio of the 1958 to 1959 differences between new and old figures. It would make very little difference if almost any other reasonable method of extrapolation were used, since the differences are so insignificant before 1958.

TABLE 4—MAN-HOURS WORKED PER YEAR IN SOVIET INDUSTRY

Year	Industrial Workers and Personnel (millions)	Average Number of Days Worked Per Year	Average Length of Working Day (hours)	Industrial Man-Hours Per Year (billions)
1950	15.324	276.3	7.96	33.703
1951	16.241	275.7	7.96	35.642
1952	16.889	274.8	7.96	36.943
1953	17.641	274.5	7.96	38.546
1954	18.535	273.9	7.96	40.411
1955	18.922	273.3	7.96	41.164
1956	19.641	272.1	7.6	40.617
1957	20.312	267.4	7.5	40.736
1958	20.988	268.0	7.433	41.809
1959	21.670	266.5	7.35	42.447
1960	22.620	266.9	6.985	42.170
1961	23.820	264.2	6.67	41.976
1962	24.677	263.4	6.67	43.355
1963	25.442	264.5	6.67	44.885
1964	26.313	266.5	6.67	46.773
1965	27.447	266.4	6.67	48.771
1966	28.514	263.1	6.67	50.039
1967	29.448			51.678
1968	30.428			53.397

The second column of Table 4, average number of days worked per year, is from the Promyshlyennost section of *Narkhoz* under the title "Ispolzovanie Kalendarnovo Vryemyeni Rabochikh v Promyshlyennosti."

Column three lists the average scheduled number of hours worked per day per wage worker in Soviet large scale industry. From 1950 to 1955 the average work day was 7.96 hours (*40 Years of Soviet Power*, p. 296; *Strana Sovietov za 50 Let*, p. 223). In February of 1956 the Twentieth Congress of the C.P.S.U. called for a program to gradually reduce the standard work week from 48 to 41 hours. In March of 1956, hours worked on Saturdays and before special holidays were reduced from 8 to 6. Thereafter the changes were somewhat more gradual. The mid-1956 effective working day was 7.6 hours (*40 Years of Soviet Power*, p. 296); mid-1957 was 7.5 hours (57 *Tsifrakh*, p. 420); end of year 1958 was 7.4 hours (58 *Narkhoz*, p. 665); end of year 1959 was 7.3 hours (59 *Narkhoz*, p. 596); end of year 1960 was 6.67 hours (60 *Narkhoz*, p. 645). Middle of the year figures from 1961 on are steady at 6.67 hours. Middle of the year figures from 1958 to 1960 are obtained as a linear combination of figures from the two nearest preceding and following dates.

In 1967, changeover to the five day working week began in earnest. Since the length of the work week was intended to be about equal before and after the change, I have assumed that the 1966 hours worked per worker per year applied to years after 1966.

Strictly speaking, the average established length of the working day applies to large scale industry only. Using it as a proxy for all industry is probably fairly safe, especially after 1959 when the producers' cooperatives were abolished. But its utilization in composing the labor series of Table 2 is far more conjectural since it is known, e.g., that work week reductions in nonindustrial sectors lagged behind their introduction in industry.

In using hours worked as an index of labor inputs, we are in effect operating with a hidden assumption that labor efficiency

per hour is unaffected by the reduction over time of total hours worked. To the extent that this is untrue, the labor index presumably ought to grow more rapidly in the period 1956–60.

I have chosen to work with unadulterated man-hours as an index of labor inputs without attempting to correct for quality differences between economic sectors or job classifications. This is primarily because it is difficult to find a good indicator of quality differences. Using wage rates to mirror marginal productivities does not strike me as appropriate to the Soviet scenario where, e.g., coal miners and fishermen make almost twice as much as the average industrial worker not because they are more "productive," but because their work is considered more dangerous and undesirable. One could try to measure and include so-called "human capital," which has undoubtedly grown very rapidly along with educational attainment in the U.S.S.R., but any estimates would have to contain a powerful amount of subjective judgment. I am left with raw man-hours not because I believe in this measure for any good a priori reasons, but simply because the alternatives seem less viable from a practical standpoint.

Capital

The industrial fixed capital stock index of Table 1 is based directly on official series. Soviet "industrial productive basic funds" appear in most official statistical books as a July 1, 1955 constant price series expressed as an end of the year index with the end of 1940 equal to 100. The series measures capital stock gross of depreciation existing at the end of a given year. It is reproduced in Table 5 along with a citation of origin. This is not the place to go into details about how the capital estimates are made by the Soviets. There is still a great deal that is unknown about the process. However, it is reasonably clear that the Soviets could, if they wanted to, put together a good capital stock series. Data on constant price purchases and retirements of capital are known to be centrally collected from each industrial enterprise. In addition, an exhaustive capital stock inventory of industrial organizations (and others) was undertaken at the end of 1959. Neither of these kinds of information are available, e.g., to estimators of American industrial capital stock.

TABLE 5—OFFICIAL SOVIET END OF YEAR CONSTANT PRICE INDUSTRIAL PRODUCTIVE BASIC FUNDS (1940 = 100)

Year	Industrial Productive Basic Funds	Source
1949	130	64 *Promyshlennost*
1950	141	67 *Narkhoz*
1951	161	64 *Promyshlennost*
1952	179	64 *Promyshlennost*
1953	198	64 *Promyshlennost*
1954	221	64 *Promyshlennost*
1955	250	67 *Strana Sovietov*
1956	277	64 *Promyshlennost*
1957	305	64 *Promyshlennost*
1958	341	65 *Narkhoz*
1959	379	64 *Promyshlennost*
1960	424	68 *Narkhoz*
1961	472	62 *Tsifrakh*
1962	523	63 *Narkhoz*
1963	586	64 *Narkhoz*
1964	653	65 *Narkhoz*
1965	715	68 *Narkhoz*
1966	779	68 *Narkhoz*
1967	840	68 *Narkhoz*
1968	907	68 *Narkhoz*

It has been noted by a few Soviet economists that official reporting of real investment may have been overstated by varying amounts (especially since 1960) because of hidden price increases.[22] Then there are the usual difficulties with new types of capital—if these are introduced at too high prices, as is sometimes alleged, there will also be a tendency to inflate growth rates of capital stock. Without inside knowledge of the situation it is difficult to evaluate which claims or counter-claims are relevant.

On the other side of the coin, to the extent that one really wants capital *services* (with imputed rentals as weights) as opposed to capital *stocks* (with capital prices as weights), the growth rates of capital stock presented in Table 1 are biased downwards. With the amount of shorter lived equipment increas-

[22] See, e.g., V. P. Krasovsky.

TABLE 6—DERIVATION OF 1960 VALUE-ADDED WEIGHTS FOR INDUSTRY

Sector	Number of Employees (thousands)	Average Yearly Earnings (roubles)	Social Insurance Deductions as percent of Earnings	Total Payments to Labor (billion roubles)	Mid-1960 Fixed Capital (billion roubles)	Interest plus Amortization Charges (Percent)	Imputed Gross Returns to Capital (billion roubles)	Total Imputed Factor Payments (billion roubles)
Electricity	339.6	1124.4	6.3	.406	11.610	12.6	1.463	1.869
Fuels	1557.2	1849	8.9	3.131	14.215	16.59	2.358	5.489
Ferrous Metals	1047.2	1401.6	7.9	1.584	8.892	13.51	1.201	2.785
Nonferrous Metals	464	1600	7.9	.801	4.276	13.51	.578	1.379
Wood and Paper	2597.5	1040.4	4.7	2.829	5.278	15.83	.836	3.665
Construction Materials	1493.4	1027.2	6.1	1.628	5.279	14.45	.763	2.390
Chemicals	739	1167.6	8.4	.935	5.084	13.54	.688	1.624
Glass and Ceramics	226.3	1027.2	6.1	.247	.418	13.8	.058	.304
Machine Building and Metal Working	7064.6	1119.6	7.5	8.503	16.505	13.71	2.263	10.766
Soft Goods	3893	790.8	6.8	3.288	4.174	13.6	.568	3.856
Processed Foods	2146	877.2	6.8	2.011	8.224	13.64	1.122	3.132

ing over time relative to longer lived structures, an ideal index of capital services would presumably grow faster than the one presented here.

In the capital series as presented, no adjustment was made for capital utilization, a standard headache in most production function estimates. The reduction in the length of the work week has probably cut back the "capital hours" input in noncontinuous process industries over the period 1956–60, but by how much is difficult to say since data on utilization rates are not available. Our disregard of this factor is equivalent to assuming that the same capital stock could be spread out evenly over the new working hours, which is probably true of capital intensive processes since they are usually multi-shifted to begin with. At the opposite extreme would be the assumption that capital hours are diminished by the same percentage as the working hours reduction. Casual playing with numbers suggests that since the changes are limited to noncapital intensive processes for the years 1956–60, overall differences in capital estimates are not significant. To some extent the fact that we have not credited the possible increased efficiency per hour of labor working fewer hours per week tends to offset possible overestimation of capital inputs due to unaccounted reductions in the length of the work week.

Let $J(t)$ be the capital stock at the *end* of year t. Exponential interpolation is performed to obtain the average capital stock during year t. The average capital stock during year t, $K(t)$, is $J(t-1)\int_0^1 e^{\gamma t}dt$ where $e^\gamma = J(t)/J(t-1)$. This yields the interpolation formula $K(t) = (J(t)-J(t-1))/(\log J(t) - \log J(t-1))$ which is the figure indexed in Table 1 for year t, except for normalization to 1960 = 100.

Output

With the exception of nonferrous metals in years past 1959, the sources of the industrial output index of Table 1 are official Soviet sectoral gross value of output series. These are aggregated with synthetic 1960 value-added weights to form an index of industrial production. The sectoral series are displayed in Table 7 normalized to 1960 = 100.

The underlying sectoral indexes are composed by the Soviets for each sector by adding together physical output units weighted by constant wholesale prices. From 1950 to 1955, January 1, 1952 prices are used; from 1955 on, July 1, 1955 prices. Our industrial production index is thus a "hybrid" mixture of gross value of output sec-

TABLE 7—SOVIET INDUSTRIAL PRODUCTION (DIRECT SOVIET SOURCES, 1960=100)

	Weights[a]	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
Electricity	5.01	27.58	31.71	35.69	41.15	46.61	53.98	60.77	67.70	77.43	89.23
Fuels	14.73	41.08	45.04	49.58	53.82	58.92	66.57	71.39	77.62	86.12	92.35
Ferrous Metals	7.47	37.03	42.68	48.12	53.56	59.00	65.48	71.55	76.36	82.43	90.38
Nonferrous Metals	3.70	33.88	39.25	46.42	51.47	58.79	65.96	71.01	76.06	84.36	91.69
Wood and Paper Products	9.84	46.60	52.45	55.44	57.82	64.39	69.39	72.96	78.57	86.73	95.24
Construction Materials	6.42	18.51	23.32	27.16	31.06	35.99	42.95	49.18	59.07	71.22	84.92
Chemicals	4.36	25.42	30.87	35.41	40.86	46.95	56.29	63.81	71.73	80.93	89.75
Glass and Ceramics	.82	27.40	31.15	35.19	39.36	45.76	53.69	61.34	68.85	77.61	88.73
Machine Building and Metal Working	28.89	23.81	27.91	32.34	37.43	43.52	51.61	58.69	66.45	75.64	86.93
Soft Goods	10.35	40.00	47.86	51.79	57.14	65.36	71.07	76.43	80.71	86.79	93.93
Processed Foods	8.41	42.54	48.25	52.63	58.77	64.91	68.42	75.00	82.02	87.28	95.61
Index of Industrial Production	100.00	33.15	38.20	42.56	47.35	53.46	60.34	66.33	73.01	81.29	90.56
	Weights[a]	1960	1961	1962	1963	1964	1965	1966	1967	1968	
Electricity	5.01	100.00	112.30	127.29	143.66	160.03	177.88	193.66	210.77	233.04	
Fuels	14.73	100.00	105.67	111.33	119.83	128.90	137.11	145.33	154.67	161.47	
Ferrous Metals	7.47	100.00	108.37	117.57	126.78	137.03	146.86	157.74	168.83	178.45	
Nonferrous Metals	3.70	100.00	108.86	118.48	127.95	137.81	149.60	162.71	178.35	191.09	
Wood and Paper Products	9.84	100.00	104.08	109.18	115.31	122.45	127.55	131.97	141.84	149.66	
Construction Materials	6.42	100.00	111.81	120.79	129.38	140.74	153.08	167.77	184.01	198.35	
Chemicals	4.36	100.00	112.58	128.79	149.29	171.47	194.68	220.10	249.16	279.38	
Glass and Ceramics	.82	100.00	111.68	123.50	134.49	149.93	161.34	176.22	194.02	214.19	
Machine Building and Metal Working	28.89	100.00	115.06	132.56	149.72	164.01	179.51	200.78	225.47	252.27	
Soft Goods	10.35	100.00	104.29	107.50	109.64	112.86	113.57	123.93	137.50	149.64	
Processed Foods	8.41	100.00	106.58	116.23	122.37	125.88	142.54	149.12	160.53	169.74	
Index of Industrial Production	100.00	100.00	109.56	120.27	131.21	141.78	153.41	167.09	183.65	199.91	

[a] 1960 value-added weights in percent.

toral indexes weighted by sectoral value-added. There are, of course, objections to the use of a gross value of output sectoral index. Among other things, such an index need not be invariant with respect to the degree of aggregation. But under the circumstances not much in the way of a feasible alternative can be used, since even the Soviet production indexes calculated in the West are composed along similar lines.

For most sectors with standardized products, the Soviet indexes of Table 7 differ but slightly (growth rates less than 10 percent different) from the latest Western estimates based on a price weighted sample of products reported in the yearbooks in physical units.[23] I choose the Soviet indexes for

[23] See, e.g., Noren. Note that most Western estimates of Machine Building and Metal Working (but not Noren's) are based on a sample of products which does not include any military equipment. This accounts for much of the difference between the two indexes.

one reason only—their more comprehensive coverage of products. Western critics argue that growing specialization and a high rate of introduction of new products into the index on a dubious price basis overstate growth in advanced multiproduct sectors like Chemicals or Machine Building and Metal Working. The Soviets retort that it is in just these sectors that the Western "sample" of products expressed in physical units is most limited to unrepresentative standard old fashioned models or types, and point out the same discrepancies in *U.S.* production indexes.[24] In any case, the difference is not that great (taking into account that Western indexes are limited to civilian production), and the same trends are evident in both series.

The Soviet sectoral indexes are listed in terms of 1940=100 in every *Narkhoz*, *Tsifrakh*, and *Promyshlyennost* volume under the title "Tyempi Rosta Valovoi Produktsi Promyshlyennosti po Otraslyam." The rule that the figures appearing in the latest yearbook take precedence is more important here because minor revisions have occurred from time to time. These sector indexes are listed in Table 7 normalized to 1960=100.

After 1959, the Soviets stopped reporting a nonferrous metals index. The nonferrous metals series linked to 1959 was constructed by James Noren, who was kind enough to supply me with an updated copy. Noren's nonferrous metals series is a weighted index of the production of thirteen nonferrous metals aggregated in July 1, 1955 enterprise wholesale prices. Estimates of metal output are computed from reports in the daily press, and from economic and technical journals.

The derivation of the sector of origin value-added weight used in aggregating the sectoral indexes into an overall industrial production index is detailed in Table 6. Number of employees is from *Trud v SSSR* (p. 86) except for nonferrous metals employment, estimated by Noren (p. 280). Yearly earnings is largely on the basis of *Trud v SSSR* data on page 140 plus educated guesses about the situation in nonferrous metals and glass and ceramics. Social insurance deductions are based on the information in Noren (p. 304). January 1, 1961 fixed capital stock is estimated on the basis of information contained on pages 215–16 of 67 *Narkhoz* for industry as a whole (88.73 billion roubles) and for all sectors except nonferrous metals and glass and ceramics. The latter are obtained as a percentage of total industrial capital stock on January 1, 1960 based on the input-output data of Joint Economic Committee (1968, p. 147) and then multiplied by 88.73 to convert to January 1, 1961 stock. Middle of the year capital stock is interpolated on the basis of the information in 64 *Promyshlenost* (p. 69). To obtain capital charge percentages, an assumed interest rate of 10 percent is added to amortization rates (excluding capital repairs) according to the new norms introduced on January 1, 1963. Nonferrous metals amortization of 3.51 percent is assumed identical to that for ferrous metals.

The industrial production index appears to be remarkably insensitive to changes in the value-added weights caused by somewhat different assumptions used in composing them. For example, using 7 percent or 15 percent as an interest rate instead of the assumed 10 percent makes very little difference on the index of industrial production. The 10 percent figure seems about right for Soviet industry in 1960, reflecting a somewhat higher degree of capital scarcity than would be present in *U.S.* industry. But it is difficult to be dogmatic about this issue, and a different value might well be more appropriate (although any reasonable value would not materially change the index of industrial production presented in Table 1).

In pretending that industrial output is a function of capital and labor alone we are introducing an error by omitting industrial purchases of inputs from other sectors, principally transportation and agriculture. My distinct feeling is that such omissions

[24] A. P. Revenko (ch. 2) presents a very worthwhile account of comparing Soviet and *U.S.* production indexes of differing origin. He concludes that production indexes of Soviet origin have no more of an inherent upward bias over time than do the standard American production indexes.

turn out to be empirically irrelevant for industry as a whole. In any case it is difficult to obtain year to year input-output type data of the kind needed to test the importance of this omission.

Data for 1969

All data for this year are preliminary. They have been obtained from the report on fulfillment of the 1969 plan from the January 25, 1970 edition of *Pravda*, and may be subject to later revision. 1969 data were included at the last minute only to bring the Table 1 series used in this study as up to date as possible.

REFERENCES

K. J. Arrow, H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

B. Balassa, "The Dynamic Efficiency of the Soviet Economy," *Amer. Econ. Rev.*, May 1964, *54*, 490–506.

E. R. Brubaker, "Embodied Technology, the Asymptotic Behavior of Capital's Age, and Soviet Growth," Aug. 1968, *50*, 304–11.

M. Feshbach, "Manpower in the U.S.S.R.: A Survey of Recent Trends and Prospects," in Part III of Joint Economic Committee 1966, pp. 703–88.

R. J. Gordon, "Problems in the Measurement of Real Investment in the U.S. Economy," unpublished doctoral dissertation, M.I.T., June 1967.

N. M. Kaplan, "The Retardation in Soviet Growth," *Rev. Econ. Statist.*, Aug. 1968, *50*, 293–303.

V. P. Krasovsky, *Problemy Ekonomiki Kapital'nykh Vlozheniy*, Moscow 1967.

D. W. Marquardt, "An Algorithm for Least Squares Estimation of Non-Linear Parameters," *J. Soc. Ind. Appl. Math.*, June 1963, *11*, 431–41.

R. H. Moorsteen and R. P. Powell, *The Soviet Capital Stock 1928–1962*, Homewood 1966.

——— and ———, *Two Supplements to the Soviet Capital Stock*, The Economic Growth Center of Yale University, New Haven 1968.

M. Nerlove, "Recent Empirical Studies of the CES and Related Production Functions," in *The Theory and Empirical Analysis of Production*, National Bureau of Economic Research, New York 1967.

R. R. Nelson, "The CES Production Function and Economic Growth Projections," *Rev. Econ. Statist.*, Aug. 1965, *47*, 326–28.

J. H. Noren, "Soviet Industry Trends in Output, Inputs, and Productivity," in Part II-A of Joint Economic Committee, Washington 1966, pp. 271–326.

A. Nove, *The Soviet Economy*, 2nd ed., New York 1969.

A. P. Revenko, *Sopostavlenie Pokazatelei Promyshlenovo Proizvodstva S.S.S.R. i S.Sh.A.*, Moscow 1966.

M. L. Weitzman, "Soviet Postwar Economic Growth and Capital Labor Substitution," Cowles Foundation Discussion Paper No. 256, 1968.

Tsentral'noe Statisticheskoe Upravlenie pri Sovete Ministrov S.S.S.R., *Forty Years of Soviet Power* Moscow 1958.

———, *Narodnoe Khoziaistvo S.S.S.R.*, Moscow 1956–1968.

———, *Promyshlenost S.S.S.R.*, Moscow 1957, 1964.

———, *S.S.S.R. v Tsifrakh*, Moscow 1959–1967.

———, *Strana Sovetov za 50 Let*, Moscow 1967.

———, *Trud v S.S.S.R.*, Moscow 1968.

U.S. Congress, Joint Economic Committee, *New Directions in the Soviet Economy*, Washington 1966.

———, *Soviet Economic Performance: 1966–67*, Washington 1968.

The Peak Load Problem with Increasing Returns and Pricing Constraints

By HERBERT MOHRING*

Transportation, electricity generation and distribution, and most other activities that are commonly classed as "public utilities" have two distinguishing characteristics that seem at least partially responsible for their being dealt with as separate topics of economic analysis: scale economies and the existence of a peak load problem. As for the former attribute, there is ample evidence that increasing returns characterizes, *inter alia*, electric power generation, the distribution of electricity and natural gas, and the provision or utilization of rights of way, rolling stock, or termini for many forms of transportation.

As for the peak load problem, the demand for the services rendered by most public utilities exhibits substantial variations from hour to hour, day to day, or season to season—periods of time too short to permit capital stock to be varied so as to keep price continuously equal to long-run marginal cost. Furthermore, the costs of storing these services are so high that the possibility of meeting short-run demand shifts by varying inventory holdings can be ignored. More accurately, at most small errors are introduced by assuming that a public utility service consumed at any instant of time must be produced at that instant. Individual demand periods, then, can be regarded as linked to each other only through their common utilization of a fixed physical plant.

The first section of this paper demonstrates that the rules for efficient allocation of resources to the production of peak load commodities are essentially the same as those for commodities which are not subject to short-run variations in demand: For both sorts of commodities, (a) price equals short-run marginal cost is a necessary condition for the efficient utilization of any given level of fixed capital plant; (b) an efficient level of plant, in turn, can be characterized by equality at the margin of the quasi-rents generated over the course of a demand cycle by a unit of capital plant with the cost per demand cycle of that unit; and (c) simultaneous satisfaction of both these equalities will yield total revenues just sufficient to cover total costs only if the production process is characterized by constant returns to scale.

For whatever the reasons, these rules for efficient resource allocation are often violated in the pricing and investment policies followed in the provision of peak load commodities. Prices generally do not vary over the course of the demand cycle to the degree that marginal costs vary. Revenues are normally required to cover total costs despite the presence of scale economies. The second section develops and interprets pricing and investment rules of relevance when there exist constraints on pricing of the sort found in the production of many transportation and

* Professor of economics at the University of Minnesota. I am deeply indebted to the Ford and the National Science Foundations for the financial support that made this research possible and to Thomas Pinfold both for helpful suggestions and for checking algebra. The approach I have taken owes a great intellectual debt to articles by Robert Strotz and Marcel Boiteux, the latter as interpreted by Jacques H. Dreze in an article and during a seminar at the University of Minnesota several years ago.

public utilities services.[1] In both sections of the paper, attention is restricted to a peak load commodity in which only two time periods—peak and off-peak—are of relevance. However, the analysis is set up in a fashion that permits straightforward generalization to any number of time periods.

I. *The Peak Load Problem Without Pricing Constraints*

Most prior analyses of the peak load problem (see, for example, articles by Peter O. Steiner and Oliver E. Williamson), have assumed as an objective the maximization of an expression like:

$$\sum_t \int_0^{q_t} f^t(q)\, dq - C(q_1, \ldots, q_n)$$

where $f^t(q)$ is the inverse demand function for the peak load commodity during period t, and C is the total cost of producing $q_1, \ldots, q_n$ units of the commodity during the respective demand periods.[2] Implicit in an objective function of this type are quite restrictive assumptions about both consumer tastes and the social welfare function. The empirical validity of an assumption that the demand for the peak load commodity in one period is independent of its price in any other period seems questionable. Furthermore, for the demand price of a commodity to depend only on its output level requires that utility functions (perhaps after suitable transformation) can be written in the form

$$u = X_1 + \sum_{j=2}^{n} f^j(X_j)$$

(See Paul A. Samuelson 1942, pp. 84–86.)

An objective function of the type that Steiner and Williamson employ says that, regardless of relative prices and of the distribution of income, society values equally a dollar received by any of its members. Under this circumstance, if social welfare is regarded as a function of individual utility levels, this function must take on the form $W = \sum_i u^i/\eta^i$ where u^i is individual i's utility index and η^i is his marginal utility of income.[3]

Conclusions similar to those of Steiner and Williamson can be reached without these restrictive implicit assumptions. Eliminating these restrictions is not without cost, however. With a more general formulation of the social welfare function, movement toward a more efficient allocation of resources would necessarily be a good thing only if accompanied by a suitable redistribution of income. It is therefore assumed in what follows that the public authority—henceforth the government—responsible for providing the public utility service not only controls utility price and output levels but also levies head taxes to redistribute income as well as to cover whatever deficits (or to redistribute whatever surpluses) may arise in its production.

The essence of the peak load problem can most easily be dealt with by considering an economy in which three commodities are produced: X_1, the amount of public utility service produced in the peak demand period; X_2, the amount of the

[1] These constrained optimization rules form the basis for the empirical work in a forthcoming paper, "The Welfare Costs of Non-Optimal Pricing and Investment Policies in Highway Transportation," by Marvin Kraus, Thomas Pinfold, and myself.

[2] Steiner and Williamson restrict their attention to the case in which short-run marginal and average variable costs are equal for outputs less than capacity and in which short-run marginal cost is infinite at capacity output. At system capacity in this case, "short-run marginal cost" is defined to equal whatever price is necessary to equate demand with capacity output.

In an appendix, Steiner does deal with a case in which demand price in period t is a function of quantities produced in all periods. However, he succeeds in getting the correct result—that price in period t should equal marginal cost during period t—only by assuming, illegitimately, consumer's surplus maximization to be the equivalent of utility maximization.

[3] See Appendix II.

service produced during the off-peak period; and X_3, everything else. The economy has I utility maximizing consumers. Consumer $i(i=1,\ldots,I)$ owns a stock of ρ^i units of a general purpose resource. This stock yields a flow of r^i units (henceforth dollars) of resource services during each demand cycle (assumed, for specificity, to be a year in length). Regardless of public utility output levels, a dollar of resource services can be converted into one unit of commodity 3.[4] Resource services are also used as the variable input in the peak load commodity's production process. $C(X_j/\alpha_j, K) \equiv C_j$ for $j=1, 2$ would be the annual cost of these variable inputs if the service was produced at a constant rate of X_j/α_j per year. In this expression, α_1 is the fraction of the year during which peak period demand conditions are in effect, $\alpha_2=1-\alpha_1$, and K is the *annual* cost of the public utility's capital plant—the number of units of commodity 3 foregone when κ units of resources are invested in that plant.

Each consumer possesses a utility function of the form $u^i=u^i(x_1^i, x_2^i, x_3^i)$ where $\sum_i x_j^i=X_j$ and an income of r^i-h^i where h^i is whatever (positive or negative) head tax may be imposed on him. The government desires to maximize welfare—a function, $W\equiv W(u^1,\ldots,u^I)$, of individual consumer utility functions. In doing so, it can control $I+3$ variables—the I head taxes, the prices (P_1 and P_2) of the public utility service, and K, the *annual* cost of the public utility service's physical plant.[5]

[4] These assumptions can be interpreted as converting the analysis into one of partial rather than of general equilibrium.

[5] It should be emphasized that throughout this paper, terms such as "resource services" and "the marginal cost of a unit of capital plant" refer to flows, not stocks. That is, the latter term, for example, refers to the price per time period of using the services of an additional unit of capital plant, not to the market value of the resources which comprise that plant.

The resource allocation problem facing the economy can be written:

$$(1)\quad \max Z\equiv W(u^1,\ldots,u^I) + \lambda[R-\alpha_1C_1-\alpha_2C_2-K-X_3]$$

where $R\equiv \sum r^i$.

Differentiating equation (1) with respect to K yields:

$$(2)\quad \partial Z/\partial K = -\lambda(\alpha_1C_{1K}+\alpha_2C_{2K}+1)=0$$

This relationship says that public utility capacity should be adjusted so that a \$1 increase in its annual cost would result in an aggregate saving of \$1 a year in variable costs.

On (a) differentiating equation (1) with respect to the remaining $I+2$ variables over which the government has control—P_1, P_2, and the I head taxes; (b) substituting into the resulting expressions the first-order conditions for utility maximization and relationships easily derivable from them; and (c) appropriately combining the derivatives with respect to head taxes and prices: an equation system results which can be written:[6]

$$(3)\quad \begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} P_1 - C_{11} \\ P_2 - C_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

In this system, C_{jj} is short-run marginal cost during period j and $S_{jk}=\sum_i s_{jk}^i$. In turn, s_{jk}^i equals $x_{jk}^i-x_k^ix_{jh}^i$ where $x_{jk}^i=\partial x_j^i/\partial P_k$, and $x_{jh}^i=\partial x_j^i/\partial h^i$, the rate at which individual i's consumption of commodity k changes when the head tax imposed on him is altered. Since a \$1 reduction in head tax is equivalent to a \$1 increase in income, $x_{jk}^i-x_k^ix_{jh}^i$ is simply the Hicks-Slutsky pure substitution effect—the rate at which individual i's consumption of commodity j changes with changes in the price of commodity k when his income is simultaneously adjusted so as to leave him on the same indifference surface.

[6] The details of these computations are given in Appendix I.

The second-order conditions for utility maximization guarantee that the determinant of the system, $S=[S_{11}S_{22}-S_{12}S_{21}]$ is greater than zero and hence that the matrix on the left of (3) has an inverse. Thus, the not very surprising conclusion results that $P_1=C_{11}$ and $P_2=C_{22}$—price equals short-run marginal cost in each demand period—is a necessary condition for welfare maximization.

Generalization of these conclusions to any finite number of time periods is quite straightforward. If the year is divided into n time periods, for example, the counterparts of equations (2) and (3) can easily be shown to be:

$$(2') \qquad \sum_i \alpha_i C_{iK} + 1 = 0$$

$$(3') \qquad [S]\{P - C\} = \{0\}$$

where $[S]$ is an $n\times n$ matrix of aggregated pure substitution terms and $\{P-C\}$ is a vector with elements P_i-C_{ii} for $i=1, \ldots, n$.

Before moving on to consider the peak load problem when pricing constraints are present, there remains an important question: What is the relationship between the revenues and costs of the public utility service when it is provided so as to maximize welfare? If the public utility service production function is homogeneous of order β in capital service and variable inputs, it can fairly easily be shown[7] that the variable cost function, C_j, can be written in the form

$$(4) \quad C_j=C(X_j/\alpha_j, K)=KG(X_j/\alpha_jK^{\beta})=KG_j$$

Substituting the right of equation (4) for C_j in the total annual cost of providing the public utility service, $TC=\alpha_1C_1+\alpha_2C_2+K$, differentiating with respect to X_j to obtain an expression for short-run marginal cost and multiplying this expression by X_j yields, for total revenues of the public utility service:

$$(5) \qquad TR = K^{1-\beta}(X_1G_1' + X_2G_2')$$

Differentiating total annual costs to obtain the equivalent of equation (2) and rearranging terms yields:

$$(6) \qquad TC = \beta K^{1-\beta}(X_1G_1' + X_2G_2')$$

Hence, total cost equals β times total revenue.

Thus, the relationship between revenues and costs for an optimally organized activity subject to a peak load problem is precisely the same as that for an activity in which no such problem exists: for both sorts of activities, the revenues derived from marginal cost prices will just cover total costs, if and only if, production is characterized by constant returns to scale, i.e., if $\beta=1$.

II. *The Peak Load Problem With Pricing Constraints*

In this section, optimization rules are developed for situations in which one or a combination of pricing restrictions are imposed on the governmental agency responsible for providing the public utility service.

The Deficit Constraint

Suppose that production of the public utility service involves scale economies but that the government is restricted to providing the service on a self-supporting basis or, more generally, that it is allowed to incur a deficit less than that which would be necessary to equate price and marginal cost. If the government cannot discriminate, e.g., by charging block rates, this restriction can be incorporated into the analysis of Section I by adding a constraint of the form:

[7] If the production function is homogeneous of order β in capital services and variable inputs, $\mu^{\beta}X_j/\alpha_j=f(\mu C_j, \mu K)$. Setting μ equal to $1/K$ yields $X_j/\alpha_jK^{\beta}=f(C_j/K, 1)=g(C_j/K)$. If attention is restricted to those sets of input values for which the marginal products of both C_j and K are positive and decreasing, g can be inverted to yield $C_j/K=G(X_j/\alpha_jK^{\beta})$.

$$\mu[P_1X_1 + P_2X_2 + D - \alpha_1C_1 - \alpha_2C_2 - K]$$

to the objective function given by equation (1) where D (which may or may not equal zero) is the allowable deficit. This equation says, simply enough, that revenues during the peak and off-peak period plus the allowable deficit must cover the total costs incurred in providing the service.

Differentiating the amended version of equation (1) with respect to K yields:

$$\partial Z/\partial K = -(\lambda+\mu)[\alpha_1C_{1k}+\alpha_2C_{2K}+1]=0$$

This relationship says that, just as when no deficit constraint is present, annual expenditures on capacity should be adjusted so that the total saving in variable costs resulting from a \$1 increase in expenditures on capacity should equal \$1.

Proceeding in a fashion analogous to that which led from equation (1) to equations (3) leads to an equation system that can be written as:

$$\begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix}\begin{bmatrix} C_{11}-P_1 \\ C_{22}-P_2 \end{bmatrix} = \frac{\mu}{\lambda+\mu}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Inverting the matrix of pure substitution terms yields:

$$\text{(7a)} \quad P_1 = C_{11} - \frac{\mu}{\lambda+\mu}\left[\frac{S_{22}X_1 - S_{21}X_2}{S}\right]$$

$$\text{(7b)} \quad P_2 = C_{22} - \frac{\mu}{\lambda+\mu}\left[\frac{S_{11}X_2 - S_{12}X_1}{S}\right]$$

where, to repeat, $S=S_{11}S_{22}-S_{12}S_{21}$. The LaGrangian multipliers, λ and μ, can be interpreted as the rates at which welfare changes with increases, respectively, in the resource base, R, and in the allowable deficit, D. Clearly, both the former and, at least within the range of D values of concern, the latter multipliers have positive values. Both S_{11} and S_{22} are negative while S is positive. If the public utility service in peak and off-peak periods are substitute products, $S_{12}(=S_{21})$ will have a positive value and the quantities in brackets on the right of equations (7) will have negative values. Hence, optimization subject to the deficit constraint requires that a price greater than marginal cost be charged in both periods. If S_{12} is negative, however—i.e., if commodities 1 and 2 are complements—nothing would appear to prevent the existence of combinations of values of S_{11}, S_{12}, X_1, and X_2 such that absolute value of $S_{11}X_2$ is less than that of $S_{12}X_1$. Under such circumstances optimization subject to the deficit constraint would require a price *below* marginal cost during the off-peak period.

This seemingly perverse result can be made more intuitively appealing by pointing out the similarity of the behavior of a deficit-constrained government authority to that of a profit maximizing monopolist.[8] A producer who has a monopoly of two products the demands for which are interdependent would presumably desire to maximize

$$\Pi \equiv P_1X_1(P_1, P_2)+P_2X_2(P_1, P_2)-C(X_1, X_2)$$

Differentiating with regard to P_j gives:

$$\partial\Pi/\partial P_j = X_j + \sum_k (P_k - C_k)X_{kj} = 0 \qquad j=1, 2$$

or

$$\begin{bmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \end{bmatrix}\begin{bmatrix} C_1 - P_1 \\ C_2 - P_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

or

$$\text{(8a)} \qquad P_1 = C_1 - \frac{X_{22}X_1 - X_{21}X_2}{X_{11}X_{22} - X_{12}X_{21}}$$

$$\text{(8b)} \qquad P_2 = C_2 - \frac{X_{11}X_2 - X_{12}X_1}{X_{11}X_{22} - X_{12}X_{21}}$$

Equations (7) and (8) differ only in (a)

[8] The similarity between these two behavior patterns was pointed out by Dreze, pp. 27–35 and Boiteux, pp. 33–40 in dealing with a deficit constraint problem similar to the one under discussion.

that the effect of changes in the price of commodity k on the consumption of commodity j are income compensated in the former but not the latter and (b) that equations (7) include the ratio of LaGrangian multipliers, $\mu/(\lambda+\mu)$. If the monopolist's two commodities are complements—commodity two is Polaroid cameras and commodity one is Polaroid film, say—it is quite conceivable that X_{11}, X_{12}, X_1, and X_2 would have values such that $|X_{12}X_1|>|X_{11}X_2|$ and hence that profit maximization would dictate setting the price of cameras below their marginal cost.[9]

The Single Price Constraint

Suppose that both a deficit constraint and a constraint that the same price be charged in peak and off-peak periods is imposed on the government. This latter constraint could, for example, result from an electricity or gas distribution system's having available meters that are capable of measuring the amount of energy consumed, but not the time at which it is used. The effects of these two restrictions on the government's behavior can most simply be analyzed by adding a constraint of the form:

$$\mu_1(P_2-P_1) \\ +\mu_2(X_1P_1+X_2P_2+D-\alpha_1C_1-\alpha_2C_2-K)$$

to equation (1). Differentiating with respect to K yields:

$$(9)\ \partial Z/\partial K=-(\lambda+\mu_2)[\alpha_1C_{1K}+\alpha_2C_{2K}+1]=0$$

As with both of the preceding cases analyzed, this equation says that capacity should be expanded to a level such that the saving in variable costs resulting from a \$1 increase in the cost of capacity equals \$1.

Proceeding along lines similar to those which led to equations (3) and (7) yields:

$$\begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix}\begin{bmatrix} C_{11}-P_1 \\ C_{22}-P_2 \end{bmatrix} = \left(\frac{1}{\lambda+\mu_2}\right)\begin{bmatrix} \mu_2X_1-\mu_1 \\ \mu_2X_2+\mu_1 \end{bmatrix}$$

Multiplying both sides of this equation system by $[S]^{-1}$, substituting P for P_1 and P_2, and dividing the resulting two equations yields:

$$(10)\quad \frac{C_{11}-P}{C_{22}-P} = \frac{\mu_2(S_{22}X_1-S_{21}X_2)-\mu_1(S_{22}+S_{21})}{\mu_2(S_{11}X_2-S_{12}X_1)-\mu_1(S_{11}+S_{12})}$$

If the government did not operate under a deficit constraint in providing the public utility service, μ_2 would equal zero, and equation (10) would reduce to

$$\frac{C_{11}-P}{C_{22}-P}=-\frac{S_{22}+S_{21}}{S_{11}+S_{12}}$$

or

$$(11)\quad P=\sum_k(S_{k1}+S_{k2})C_{kk}/\sum_{j,k}S_{jk}\quad j,k=1,2$$

In general, for a utility maximizing consumer, $\sum_j S_{kj}P_j=0$ where the summation is over all commodities (see Samuelson 1948, p. 105). Hence, in the case presently under consideration,

[9] As Baumol and Bradford point out, if $X_{12}=X_{21}=S_{12}=0$ in equations (7) and (8), dividing them through by P_1 or P_2 as appropriate and rearranging terms yields essentially the same relationships as those first developed by Frank Ramsey:

$$(7')\quad (P_k-C_{kk})/P_k=-[\mu/(\lambda+\mu)](X_k/S_{kk}P_k)$$

and

$$(8')\qquad (P_k-C_k)/P_k=-X_k/X_{kk}P_k$$

Equation (8′) says that, for a profit maximizing monopolist, "Lerner's index of monopoly power" equals the reciprocal of the elasticity of demand for the product. Equation (7′) says that maximizing welfare subject to a deficit constraint requires that Lerner's index be set equal to the reciprocal of the (income compensated) elasticity of demand for the product times $\mu/(\lambda+\mu)$—a number that clearly lies between zero and one within the range of allowable deficit values that is of interest.

$$S_{k3} = -(S_{k1} + S_{k2})P$$

Making use of the fact that $S_{jk}=S_{kj}$ and successively substituting this relationship into equation (11) yields

$$(12) \qquad P = (S_{13}C_{11} + S_{23}C_{22})/(S_{13} + S_{23})$$

or

$$(13) \qquad S_{13}C_{11} + S_{23}C_{22} + S_{33} = 0$$

as equivalent conditions to be satisfied if the government is subject only to the constraint that the same price must be charged in both demand periods for the public utility service.

It seems worth noting that simultaneous satisfaction of equations (9) and (13) would not, in general, guarantee that revenues would equal costs for the public utility service even if it is characterized by constant returns to scale. Define S (for surplus) such that

$$P(X_1 + X_2) = S + \alpha_1 C_1 + \alpha_2 C_2 + K$$

If provision of the public utility service involves constant returns to scale, Euler's theorem permits rewriting this equation as:

$$(14) \quad P(X_1 + X_2) = S + X_1C_{11} + X_2C_{22} + K[\alpha_1 C_{1K} + \alpha_2 C_{2K} + 1]$$

Satisfaction of equation (9) guarantees that the bracketed expression on the right of equation (14) equals 0. Hence, substituting equation (12) for P in equation (14) yields, on rearranging terms

$$(15) \quad \begin{aligned} S &= (C_{11} - C_{22}) \\ &\quad \cdot (X_2S_{13} - X_1S_{23})/(S_{13} + S_{23}) \\ &= (C_{22} - C_{11}) \\ &\quad \cdot (X_2S_{13} - X_1S_{23})P/S_{33} \end{aligned}$$

Both S_{33} and, by definition, $(C_{22}-C_{11})$ are negative. Hence, given constant returns to scale and a constrained benefit maximizing price and capacity level, the public utility service will generate a surplus if, and only if, X_2S_{13} is greater than X_1S_{23}.

The Single Toll Constraint

In taking trips or shipping goods, the buyer of transportation services plays both a consuming and a producing role. At the very least, he supplies one scarce resource—his own time or that of the goods he ships—essential to performing the service. Indeed, a private passenger vehicle operator supplies not only his own time but essentially all of the remaining variable inputs involved in his travel. The required quantity of at least some of these inputs depends on the level of traffic on the road he uses. In particular, an increase in the volume of travel on a road results in an increase in the time required for a trip. In contrast, gasoline consumption appears to be independent of the rate at which freeways and rural roads are utilized but does increase with the level of traffic on urban streets.

The price an auto traveler pays for his trips, then, can be regarded as the cost of the inputs he himself provides plus whatever tax or toll is imposed on him. In the United States, the most important highway user tax is that on gasoline. To repeat, at least on freeways and rural roads, gasoline consumption and hence the tax per trip appear to be independent of the rate at which trips take place. Thus, in relying mainly on a tax on the gasoline used by freeway travelers, the government is, in effect, setting a price in each demand period of $P_j = T + \alpha_j C(X_j/\alpha_j, K)/X_j$, where T is the fixed toll and the remaining notation is that of preceding sections.

The effect of imposing this sort of constraint on the government's pricing behavior can most easily be analyzed by adding

$$(16) \quad \begin{aligned} &\mu_1[G+(P_2-\alpha_2C_2/X_2)-(P_1-\alpha_1C_1/X_1)] \\ &+\mu_2[P_1X_1+P_2X_2+D-\alpha_1C_1-\alpha_2C_2-K] \end{aligned}$$

to equation (1). The parameter, G, is introduced to allow for the possibility that the toll imposed on trips may increase with increases in the rate at which a given road is used to produce them. Once more, the real world analogue of G appears to be very close to zero on freeways and rural roads but positive, although less than optimally so, on urban streets.

Manipulating the derivatives of the amended equation (1) with respect to prices and head taxes yields a pair of equations that can be written as (17) where Y_j denotes $(C_{jj}-\alpha_j C_j/X_j)$, the difference between short-run marginal and average variable costs in period j. Multiplying both sides of equation (17) by $[S]^{-1}$, dividing the resulting two equations, and substituting in the pricing constraint,

$$T = P_1 - \alpha_1 C_1/X_1 = G + P_2 - \alpha_2 C_2/X_2,$$

yields equation (18).

Each of the expressions in brackets in equation (18) will have negative values if $S_{12}=S_{21}=0$ and if, as would normally be the case, Y_1 and Y_2 are positive. That this is the case reflects the fact that S and S_{jj} are, respectively, positive and negative for utility maximizing consumers. To repeat, within the relevant range of deficit constraint values, the LaGrangian multiplier, μ_2, is positive. The LaGrangian, μ_1, can be interpreted as the rate at which welfare changes with changes in the gap, G, between off-peak and peak period tolls. This multiplier will also be positive if the difference between these two tolls is less than the difference, Y_1-Y_2, necessary to equate price with marginal cost in each period. If both multipliers are positive and the bracketed expressions are all negative, the denominator of the expression on the right-hand side of equation (18) will definitely have a negative value while the numerator could be of either sign. *Ceteris paribus*, the greater the value of μ_2 (i.e., the more stringent the deficit constraint), the more likely it is that the numerator will also be negative and hence that a toll will be required yielding prices in excess of short-run marginal costs in both demand periods. If, however, μ_2 is sufficiently small (e.g., if no deficit constraint is imposed) that the numerator and denominator on the right of equation (18) are of opposite signs, optimization would require a toll falling somewhere between Y_1 and Y_2. That is, optimization would require a price that is simultaneously somewhat less than peak period marginal cost and somewhat greater than off-peak marginal cost.

Differentiating with respect to K, equation (1) as amended by the single toll and deficit constraints yields, after rearranging terms,

$$(19)\qquad \begin{aligned} &1 + \alpha_1 C_{1K} + \alpha_2 C_{2K} \\ &= (\alpha_1 C_{1K}/X_1 - \alpha_2 C_{2K}/X_2)\mu_1/(\lambda + \mu_2) \end{aligned}$$

In each of the preceding cases discussed, optimization required adjusting expenditures on capacity to the point where the saving in variable costs effected by spending an additional dollar on capacity equalled \$1.

$$(17)\qquad \begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} (\lambda + \mu_2)(C_{11} - P_1) - \mu_1 Y_1/X_1 \\ (\lambda + \mu_2)(C_{22} - P_2) + \mu_1 Y_2/X_2 \end{bmatrix} = \begin{bmatrix} \mu_2 X_1 - \mu_1 \\ \mu_2 X_2 + \mu_1 \end{bmatrix}$$

$$(18)\qquad \frac{C_{11} - P_1}{C_{22} - P_2} = \frac{T - Y_1}{T - G - Y_2} = \frac{\mu_2[S_{22}X_1 - S_{21}X_2] - \mu_1[S_{22} + S_{21} - SY_1/X_1]}{\mu_2[S_{11}X_2 - S_{12}X_1] + \mu_1[S_{11} + S_{12} - SY_2/X_2]}$$

With a single toll constraint (or a constraint that the difference in tolls be less than $Y_1 - Y_2$), however, the right-hand side of equation (19) would normally be negative. That is, the optimum level of capacity would be "inefficiently small"—a level for which more than a dollar of variable costs would be saved by spending an additional dollar on capacity. That this is the case reflects two facts: a) that each of the multipliers on the right of equation (19) can be expected to be positive within the relevant ranges of D and G values; and b) that the law of diminishing marginal productivity suggests that $|\alpha_1 C_{1K}/X_1|$ is normally greater than $|\alpha_2 C_{2K}/X_2|$—i.e., that the reduction in average variable costs occasioned by a given increase in capacity is an increasing function of output.

To put it differently, with a single toll constraint, if a given dose of capacity reduces peak period average variable costs by more than it reduces off-peak average variable costs, it also reduces the already too small differential between peak and off-peak prices. Under such circumstances, a movement toward the capacity level required for least cost production of the equilibrium numbers of trips would, to a degree, alter trip equilibria in an inefficient fashion. The right-hand side of equation (19) takes this fact into account.

The model summarized by equations (16) to (19) can be modified slightly to deal with the resource misallocation resulting when the same toll is charged for roads having different construction costs. Let $C_j(X_j, K_j)$ be the total variable costs of X_j trips per year on road j when K_j dollars per year are spent on road j capacity. Suppose, for specificity, that providing capacity on road 1 is more expensive than providing it on road 2.

In this two-road single toll case, the objective function can be written as equation (20). Although the interpretation of C_{11} and C_{22} is different, differentiating equation (20) with respect to P_1, P_2, and h^i, substituting consumer equilibrium relaships, etc. yield an expression for the optimum toll that can be written in precisely the same form as equation (18). Differentiating with respect to K_j yields, after rearranging terms:

$$\text{(21a)} \quad 1 + C_{1K} = [\mu_1/(\lambda + \mu_2)]C_{1K}/X_1$$

$$\text{(21b)} \quad 1 + C_{2K} = -[\mu_1/(\lambda + \mu_2)]C_{2K}/X_2$$

As with the preceding cases, setting the left sides of equations (21) equal to zero would result in a \$1 increase in capacity outlays yielding a \$1 decrease in variable costs. However, the right-hand side of equation (21a) is negative while that of (21b) is positive.[10] That is, for reasons quite similar to those mentioned in discussing the single toll-peak load case, if the same toll must be charged for all roads, achieving a constrained optimum would require building "inefficiently small" expensive roads and "inefficiently large" low cost roads. Thus, the putative fact that benefit/cost ratios are lower on rural than on urban portions of the Interstate System (see Anne Friedlaender) can at least partially be rationalized on (second best) efficiency grounds.

[10] C_{jK} is negative—an increase in capital outlays reduces variables costs. For the ranges of D and G values of relevance, the ratios of the LaGrangians on the right of equations (21) are positive.

$$\text{2(0)} \quad \begin{aligned} \text{Max } Z \equiv W &+ \lambda(R - C_1 - C_2 - K_1 - K_2 - X_2) \\ &+ \mu_1(G + P_2 - C_2/X_2 - (P_1 - C_1/X_1)) \\ &+ \mu_2(P_1X_1 + P_2X_2 + D - C_1 - C_2 - K_1 - K_2) \end{aligned}$$

Appendix I

The Derivation of Equations (3)

Consumer i $[i=1, \ldots, I]$ seeks to maximize

$$s \equiv u^i(x_1^i, x_2^i, x_3^i) + \eta^i[r^i - h^i - P_1 x_1^i - P_2 x_2^i - x_3^i]$$

The first-order conditions for this maximization problem are:

$$(22a) \qquad u_j^i - \eta^i P_j = 0 \qquad j = 1, 2$$

$$(22b) \qquad u_3^i - \eta^i = 0$$

$$(22c) \qquad r^i - h^i - P_1 x_1^i - P_2 x_2^i - x_3^i = 0$$

Substituting the demand equations that can be derived from these conditions into equation (22c), and differentiating with respect to the P_j and h^i gives:

$$(23a) \qquad x_j^i + P_1 x_{1j}^i + P_2 x_{2j}^i + x_{3j}^i = 0 \qquad j=1, 2$$

$$(23b) \qquad 1 + P_1 x_{1h}^i + P_2 x_{2h}^i + x_{3h}^i = 0$$

Summing equation (23a) over all I individuals gives:

$$(23c) \qquad X_j + P_1 X_{1j} + P_2 X_{2j} + X_{3j} = 0$$

The government is assumed to maximize:

$$(24) \quad Z \equiv W(u^1, \ldots, u^I) + \lambda[R - \alpha_1 C_1 - \alpha_2 C_2 - K - X_3]$$

Differentiating with respect to P_1, P_2 and the I head taxes yields equations (25a) and (25b). C_{kk} denotes $\partial C_k / \partial(X_k/\alpha_k)$—short-run marginal cost in period k—and the remaining notation is hopefully obvious.

Using equation (22a) and (22b) to eliminate the u_k^i and then substituting equation (23a) reduces the first summation on the right of equation (25a) to $\sum_i W_i \eta^i x_j^i$. Using equation (23c) to eliminate X_{3j} from the expression preceded by λ in equation (25a) yields:

$$(26a) \qquad -\sum_i W_i \eta^i x_j^i + \lambda[X_j + \sum_k (P_k - C_{kk}) X_{kj}] = 0 \qquad j, k = 1, 2$$

Similar substitutions of equations (22a), (22b), and (23b) into equations (25b) yield:

$$(26b) \qquad -W_i \eta^i + \lambda[1 + \sum_k (P_k - C_{kk}) x_{kh}^i] = 0 \qquad k = 1, 2$$

Finally, multiplying equation (26b) through by x_j^i, summing over all i individuals, and subtracting the resulting expression from equation (26a) yields:

$$(27) \qquad \lambda \sum_k (P_k - C_{kk})(X_{kj} - \sum_i x_j^i x_{kh}^i) \equiv \lambda \sum_k (P_k - C_{kk}) \sum_i (x_{kj}^i - x_j^i x_{kh}^i) = 0 \qquad j, k = 1, 2$$

Letting $s_{kj}^i \equiv x_{kj}^i - x_j^i x_{kh}^i$ and $S_{kj} \equiv \sum_i s_{kj}^i$, the system defined by equations (27) can be written:

$$\begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} P_1 - C_{11} \\ P_2 - C_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Appendix II

The Baumol-Bradford, Dixit, and Lerner Papers—Income Compensation and the "Tax Proportional to Marginal Cost" Theorem

In this paper, a general analytical framework associated with the names Ramsey, Boiteux, Dreze, et al. is applied to a specific problem. Papers by Baumol and Bradford,

---

$$(25a) \qquad \partial Z/\partial P_j = \sum_i W_i \sum_k u_k^i x_{kj}^i - \lambda[C_{11} X_{1j} + C_{22} X_{2j} + X_{3j}] = 0 \qquad j = 1, 2; \quad k = 1, 2, 3$$

$$(25b) \qquad \partial Z/\partial h^i = W_i \sum_k u_k^i x_{kh}^i - \lambda[C_{11} x_{1h}^i + C_{22} x_{2h}^i + x_{3h}^i] = 0 \qquad i = 1, \ldots, I; \quad j = 1, 2; \quad k = 1, 2, 3$$

Dixit, and Lerner in the preceding isssue of this *Review* all dealt with this general framework. Such differences as exist between my findings and those of my immediate predecessors are primarily the result of differences in our assumptions about the possibility of lump sum income redistribution. This is a matter of sufficient consequence to warrant further comment. In addition, the three earlier papers come to conflicting and, to a degree, ambiguous conclusions about the circumstances under which it would be efficient to levy taxes on commodities proportional to their marginal costs. Some of this apparent confusion can be eliminated. It is these two subjects to which this appendix is devoted.

It is implausible to suppose that a government faced with the pricing constraints I have analyzed would have the ability to effect costless lump sum income transfers. Still, if maximizing some social welfare function is taken to be the goal, income redistribution must, in general, accompany imposition of the pricing rules developed in this paper and dealt with by Baumol, et al. Without redistribution, social welfare could decline on moving from a nonoptimal state to one satisfying a constrained optimization rule or, for that matter, one in which marginal cost prices are established across the board.

An uncritical reader of the Baumol-Bradford demonstration of "The Formal Theorem" in Section III of their paper could overlook this fact and take the paper as arguing (perhaps even proving) that satisfaction of "their" pricing rules would suffice to achieve not only a constrained Pareto optimum but also a constrained welfare maximum. They begin their demonstration by defining $Z(P_1, \ldots, P_n)$, "our (unspecified) measure of consumer benefit which is to be maximized subject to the profit constraint $\pi(P_i, \ldots, P_n) = M$." On "examining more specifically the nature of the consumer benefit function," they conclude that its partial derivative with respect to the price of commodity j can be written $\partial Z/\partial P_j = -X_j$ and hence that quasi-optimality requires:

$$(28) \qquad \partial Z/\partial P_j = -X_j = \lambda \partial\pi/\partial P_j$$

where λ is the Lagrangian multiplier associated with the profit constraint.

That satisfaction of equation (28) would not, in general, guarantee a welfare maximum can be demonstrated quite easily. Suppose that the objective is to maximize

$$(29) \qquad W(U^1, \ldots, U^I) - \lambda[\pi(P_1, \ldots, P_n) - M]$$

Differentiating with respect to the price of commodity j and appropriately substituting the equivalents of equations (22a) and (23a) in Appendix I yields:

$$(30) \qquad -\sum_i W_i \eta^i x_j^i = \lambda \partial\pi/\partial P_j$$

where W_i is the change in social welfare resulting from a change in individual i's utility and η^i is his marginal utility of income. The summation in equation (30) reduces to the left side of equation (28) only if $W_i\eta^i = 1$ for all individuals, i.e., only if each individual's welfare weight equals the reciprocal of his marginal utility of income.

In brief, in the absence of income transfers, satisfaction of equation (28) is a necessary condition for maximizing welfare only if society regards a dollar to one individual as the welfare equivalent of a dollar to any other. This is, e.g., the interpretation of legislative intent most commonly made in justifying expenditures under laws which say, in effect, "A public investment may be undertaken only if the benefits, to whomsoever they may accrue, exceed the costs, whosoever may incur them." But the assumption is still very restrictive, common though it may be.

Lerner's "*first rule* of ideal taxation" is, "Maintain the *proportionality* of price to marginal cost by having a uniform *tax rate* on all uses of resources." He indicates this rule to be applicable when either a) such shifting of resources as results from the imposition of taxes "is only to another part of the taxed sector," or b) it is "possible to tax *all* final output."

Baumol and Bradford accept the "*first rule*" when reallocation occurs only within the taxed sector. Indeed, they present an

ingenious (if unnecessarily complicated)[11] demonstration of its validity in this case. However, they argue the rule to be invalid "when all goods are taxable." In doing so, they treat "all goods" and "goods involved in market transactions" as synonyms. In their algebraic development, they follow "the sign convention that positive numbers represent net purchases of desired goods, and negative numbers, net sales" and specify relationships so that positive taxes on the goods consumers purchase are exactly offset by negative taxes on the commodities they sell. Dixit employs the same convention in his discussion of Lerner's "first rule."

The Baumol-Bradford conclusion does, indeed, follow from their premises. However, their definition of "all goods" differs fundamentally from that employed by Lerner. He clearly defines "final output" to include such cases of resources as leisure and do-it-yourself. He admits it to be "no more possible to tax *all* final outputs than to have lump sum taxes." Still, his rule would be valid if "*all* final outputs" could be taxed. That this is the case can be demonstrated by analyzing a simple but easily generalized example.

Suppose, then, that the single citizen of an economy gets utility from a commodity, X, from working L hours a day, and from enjoying recreational activities, R, during the $24-L$ hours that remain each day. For convenience, units are specified so that one hour's labor produces one unit of commodity X. The consumer's budget constraint can be written:

$$PX + t_R R = PX + t_R(24 - L) = WL$$

where $P=1+t_X$, $W=1-t_L$, and t_X, t_L, and t_R are the taxes on commodity X, labor, and recreation, respectively. Maximizing his utility function, $U(X, L, R)$, subject to this constraint yields as first-order conditions, $U_X=\lambda P$ and $U_R-U_L=\lambda[W+t_R]$. These relationships together with his budget and time constraints suffice to determine his consumption of X and allocation of time between work and recreation as functions of W, P, and t_R. Inserting these relationships into the consumer's budget constraint and differentiating with respect to t_X yields

$$(31) \quad (W + t_R)L_P - PX_P - X = 0$$

where the P subscript indicates partial differentiation. Similar relationships result on differentiating with respect to t_R and t_L.

The government of this economy desires to collect C units of commodity X from the consumer in a fashion that minimizes the burden of the tax on him. That is, it desires to select those values of t_X, t_R, and t_L which maximize:

$$U(X, L, R) + \mu[t_X X + t_R R + t_L L - C]$$

Differentiating this expression with respect to t_x, making appropriate substitutions of the consumer's first-order conditions for utility maximization and equation (31), and rearranging terms yields:

$$(32) \quad (1-\lambda/\mu)X+t_X X_P+(t_L-t_R)L_P=0$$

Again, similar relationships result on differentiating with respect to t_L and t_R. On multiplying equation (31) through by $(1-\lambda/\mu)$ and comparing the coefficients of X_P and L_P in the resulting relationship and equation (32), it can be seen that

$$t_X = (1 - \lambda/\mu)P = (1 - \lambda/\mu)(1 + t_X)$$

or,

$$t_X = \mu/\lambda - 1$$

and that:

$$t_L - t_R = -(1 - \lambda/\mu)(W + t_R)$$
$$= (1 - \lambda/\mu)(t_L - t_R - 1)$$

or:

$$t_L - t_R = (1 - \mu/\lambda) = -t_X$$

[11] They note that this alternative "is obviously equivalent to the assertion that labor is perfectly inelastically supplied over a wide range of price vectors." This being the case the efficiency of a proportionate tax would seem to follow directly from the fact that an α percent tax on all commodities sold to a wage earner who spends all his income would have the same effect on his behavior as an $\alpha/(100+\alpha)$ percent tax on his wages. Thus, a uniform excise tax would serve effectively to tax an input in completely inelastic supply. It would therefore lead to no resource misallocation.

That is, to maximize the consumer's utility while collecting C units of commodity X from him requires that the tax on a leisure hour be set equal to the sum of the taxes on a unit of commodity X and on one working hour. If only final goods can be taxed (i.e., if no tax on wages is possible), then, as Lerner's *first rule* indicates, the tax on an hour of recreation must equal that on the commodity X producible in one hour. If *all* commodities can be taxed, proportional levies are, indeed, optimal.

REFERENCES

W. J. Baumol and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *60*, 265–83.

M. Boiteux, "Sur la Gestion des Monopoles Publics Astreints a l'Equilibre Budgetaire," *Econometrica*, Jan. 1956, *24*, 22–40.

A. Dixit, "On the Optimum Structure of Commodity Taxes," *Amer. Econ. Rev.*, June 1970, *60*, 295–301.

J. H. Dreze, "Some Postwar Contributions of French Economists to Theory and Public Policy, with Special Emphasis on Problems of Resource Allocation," *Amer. Econ. Rev.*, June 1964, *54*, 1–64.

A. F. Friedlaender, *The Interstate Highway System*, Amsterdam 1965.

A. P. Lerner, "On Optimal Taxes with an Untaxable Sector," *Amer. Econ. Rev.*, June 1970, *60*, 284–94.

F. P. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.* Mar. 1927, *37*, 47–61.

P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge, Mass. 1948.

———, "Constancy of the Marginal Utility of Income," in Oscar Lange, et al., eds., *Studies in Mathematical Economics and Econometrics, in Memory of Henry Schultz*, Chicago 1942, 75–91.

P. O. Steiner, "Peak Loads and Efficient Pricing," *Quart. J. Econ.*, Nov. 1957, *71*, 585–610.

R. Strotz, "Urban Transportation Parables," in Julius Margolis, ed., *The Public Economy of Urban Communities*, Washington 1965, 127–69.

O. E. Williamson, "Peak Load Pricing and Optimal Capacity," *Amer. Econ. Rev.*, Sept. 1966, *56*, 810–27.

COMMUNICATIONS

Manpower Programs in a Local Labor Market: A Theoretical Note

By DANIEL S. HAMERMESH AND ROBERT S. GOLDFARB*

A major aspect of social policy in the United States during the 1960's was the effort to increase employment and lessen the extent of poverty. The effects of these efforts, in particular those of the Manpower Development and Training Act, have been discussed by economists solely within the framework of empirical cost-benefit analysis. In this note we take a different approach to the study of manpower programs. Under a set of admittedly restrictive assumptions we analyze the relative efficiencies in reducing unemployment of several alternative subsidy programs. It should be remembered that the reduction of unemployment is merely one goal of these programs and that the usefulness of our result must be qualified accordingly.

I. *The Dominance of Wage over Training Subsidies in the Short Run*

The closed labor market of our model contains a large number of firms which are price-takers in their respective product markets. The supply of labor is infinitely elastic at an unskilled wage rate, which for simplicity is assumed to equal one, up to some full-employment level E_f at which the supply becomes completely inelastic.[1] We assume that all labor is homogeneous until trained for a specific job, that all training is specific to the firm which undertakes it and that it takes one period to train the newly-hired worker. Furthermore, each firm maximizes its profits by choosing both a rate of output Y and a mark-up m over the unskilled wage which was assumed above to equal one. The optimal wage in each firm, $1+m$, will vary directly with the amount of training required to enable the worker to take part in the production process.

We consider a single firm in this labor market and examine its behavior first in a world without government intervention and then in response to different incentives offered by the government. Workers quit this firm at the rate:

$$Q = Q(1 + m),\ Q' < 0,\ Q'' > 0,$$

which we can condense to:

$$q = q(m),\ q' < 0,\ q'' > 0$$

The negative relation between q and m is well established (see Walter Oi, p. 545, for example), and it seems reasonable to assume that successive increases in the wage have decreasing success in reducing the quit rate. We assume that the firm's capital stock is fixed, so that our analysis is short-run. We can write the production function with only employment E as its argument:

$$Y = f(E),\ f' > 0,\ f'' < 0$$

The firm's profits in each time period are:

$$(1)\quad \pi = pf(E) - [1 + m]E - q(m)ET,$$

where p is the product price and T is the required expenditure per worker which we

* The authors are assistant professors of economics at Princeton University and Yale University, respectively. They wish to thank Mark Leiserson and Paul Gayer of Yale for helpful comments. The Ford Foundation and the Manpower Administration of the U.S. Department of Labor provided research support; views expressed in the paper are solely the authors'.

[1] We make this assumption in order to concentrate on the demand aspects of manpower subsidies. A large literature, summarized in Peter Diamond, discusses income maintenance and public assistance with this assumption relaxed, but ignores the demand for labor.

assume to be fixed in the short run.[2] The firm maximizes profits by solving the following pair of equations for the decision variables m and E.[3]

$$(2) \qquad -E[1+q'(m)T]=0,$$

and

$$(3) \qquad pf'(E)-[1+m]-q(m)T=0$$

Condition (2) states that the firm will set a wage mark-up m such that the increase in direct wage costs resulting from a slight increase in m is just offset by the decrease in replacement costs. Condition (3) states that the marginal product of labor will be set equal to the wage plus the replacement cost of training.

What happens to the wage mark-up and employment in this profit-maximizing firm when the government offers it subsidies aimed at increasing its employment? We assume three alternative subsidy schemes: Case 1, a dollar-amount wage subsidy; Case 2, a percentage wage subsidy; and Case 3, a dollar-amount subsidy for training costs. We ignore a percentage subsidy for training because it is analytically identical to the dollar-amount subsidy which we do analyze. The training subsidy in our discussion is a fair analytic description of the on-the-job training aspects of the Manpower Act. The wage subsidies, though different from any labor market scheme in operation today, are analogous to the investment tax credit in the market for real capital. We assume here that both the wage and training subsidies apply to all workers hired and employed by the firm selected to receive them.

Profits can be represented by the general form:

$$(1') \qquad \pi=pf(E)-E\{[1+m][1-x]-c\} - q(m)E[T-t],$$

where

c = dollar-amount wage subsidy,

x = percentage wage subsidy, and

t = dollar-amount training subsidy.

The general profit-maximizing conditions for the firm are:

$$(2') \qquad 1-x+q'(m)[T-t]=0,$$

and

$$(3') \qquad pf'(E)-[1+m][1-x] + c-q(m)[T-t]=0$$

The solutions for the decision variables m and E are:

$$(4) \qquad m^*=q'^{-1}\left(\frac{-[1-x]}{T-t}\right),$$

and

$$(5) \qquad E^*=f'^{-1}\left(\frac{1}{p}\,[[1+m^*][1-x] - c+q(m^*)[T-t]]\right)$$

Since the derivative of the function q'^{-1} is positive, it can be shown by differentiating in (4) and (5) that the derivatives of E^* with respect to c, x, and t are all positive and that:

$$\frac{\partial m^*}{\partial c}=0;\ \frac{\partial m^*}{\partial x}>0;\ \frac{\partial m^*}{\partial t}<0$$

The dollar-amount wage subsidy has no effect on wages because it does not change the terms on which the firm trades off between direct wage costs and turnover costs. The percentage wage subsidy decreases direct wage costs relative to turnover cost at any

[2] The assumption of fixed per-capita training costs follows logically from the short-run fixity of the capital stock. Once the firm has installed its capital equipment the type of equipment dictates the amount and kind of training to be given. In the long run the variable T is subject to choice by the firm just as is the capital stock. A long-run analysis of the problem discussed in this note might include T in the production function along with labor and capital.

[3] The second-order conditions for maximizing profits are satisfied:

$$\frac{\partial^2\pi}{\partial m^2}=-ETq''(m)<0;$$

$$\frac{\partial^2\pi}{\partial E^2}=pf''(E)<0;$$

and

$$\frac{\partial^2\pi}{\partial m\partial E}=0$$

fixed m, inducing the firm to raise its wage and lessen its total expenditures on training in each time period. The dollar-amount training subsidy produces the opposite effect, making training relatively cheaper and inducing the firm to pay a lower wage at the cost of a higher quit rate and more frequent need to incur the reduced costs of training.

For any arbitrary values of the subsidy parameters c, x, and t the optimizing values of m and E can be derived for each type of subsidy from (4) and (5). Denoting by E_i^* and m_i^* the profit-maximizing employment and wage mark-up chosen by the firm for each subsidy, we can write the total cost to the government as:

$$C_1 = cE_1^*; \tag{6}$$

$$C_2 = x[1 + m_2^*]E_2^*; \tag{7}$$

and

$$C_3 = tq(m_3^*)E_3^* \tag{8}$$

for the dollar-amount wage, percentage wage, and dollar-amount training subsidies.

Having set forth both the free-market model and the subsidy schemes to be imposed on that market, we can proceed to evaluate the relative short-run efficiencies of the three schemes by making pairwise comparisons among them. We assume that the government wishes to raise employment in this firm from the free-market E_c^* to some level $\bar{E}^*$ and then examine the cost per worker incurred by the government in increasing employment by $(\bar{E}^* - E_c^*)$.

We first compare Case 1, the dollar-amount wage subsidy, to Case 2, the percentage wage subsidy. If each scheme is to produce an increase from the free-market level of employment to an identical higher level, the argument of the function in (5) must be equal under each scheme:

$$1 + m_1^* - c + q(m_1^*)T = [1 + m_2^*][1 - x] + q(m_2^*)T,$$

or

$$c = m_1^* - m_2^*[1 - x] + x + [q(m_1^*) - q(m_2^*)]T$$

Substituting for c in (6) and comparing C_1 with C_2, we have $C_1 \gtreqless C_2$ as:

$$m_1^* - m_2^*[1 - x] + x + [q(m_1^*) - q(m_2^*)]\,T \gtreqless x[1 + m_2^*]$$

We cancel terms on both sides of the inequality and apply the mean value theorem to the difference in the quit rates to derive:

$$[m_1^* - m_2^*][1 + q'(\bar{m})T] \gtreqless 0, \qquad m_2^* > \bar{m} > m_1^* \tag{9}$$

The first term in (9) is negative, for the wage rate is increased when the firm is offered the percentage wage subsidy, but it remains unchanged for the dollar-amount subsidy. We know from (2′) that under the dollar-amount wage subisdy:

$$1 + q'(m_1^*)T = 0 \tag{10}$$

Now $|q'(\bar{m})| < |q'(m_1^*)|$ because $\bar{m} > m_1^*$ and $q'' > 0$, so that the second term of (9) must be positive. We may thus conclude that the expression on the left-hand side of (9) is negative, or that for equal employment effects, $C_1 < C_2$.[4]

We use identical methods to compare Cases 1 and 3 at a constant $\bar{E}^*$ and find that $C_1 \gtreqless C_3$ as:

$$[m_1^* - m_3^*][1 + q'(\bar{m})T] \gtreqless 0, \; m_1^* > \bar{m} > m_3^* \tag{11}$$

The first term is positive, and from (10) and a comparison of $q'(m_1^*)$ and $q'(\bar{m})$ we see that the second term is negative. Thus the expression on the left side of (11) is negative and the dollar-amount wage subsidy dominates the training subsidy, $C_1 < C_3$.

The third comparison, between the percentage wage and training subsidies, does not result in an unambiguous relation between the costs of the two programs. Using the same approach as above we find that $C_2 \gtreqless C_3$ as:

$$[m_2^* - m_3^*][1 + q'(\bar{m})T] \gtreqless 0, \; m_2^* > \bar{m} > m_3^* \tag{12}$$

[4] Throughout this note, $\bar{m}$ is used to denote any unknown wage whose value is bracketed by two solution values of m.

The sign of this expression depends on the shape of the quits function between m_2^* and m_3^*, and in particular on whether $\bar{m}$ is greater or less than the free-market m^*. If $q''(m)$ is large in the range between m_2^* and m_3^*, the value $\bar{m}$ will lie near m_3^* and will be less than the free-market m^*, so that the second term of (12) will be negative. Thus in the case of a highly convex quits function the percentage wage subsidy is more efficient. If, on the other hand, the function is only slightly convex in this range around the free-market m^*, the second term of (12) is positive and the percentage wage subsidy is less efficient than the training subsidy.

II. *Conclusion and Qualifications*

We have demonstrated that in a labor market in which workers move between firms and all training is specific to the firm, a dollar-amount wage subsidy is uniformly more efficient in increasing employment than are percentage wage or training subsidies. This result holds because among the three subsidy schemes considered only the dollar-amount wage subsidy does not alter the wage rate the firm pays. The percentage wage subsidy raises the firm's wage, and the resulting decrease in average replacement costs is less than the increase in average direct wage costs. The dollar-amount training subsidy lowers the wage, but the decrease in average direct wage costs is less than the increase in average replacement costs. Under both of these schemes the average cost of labor rises relative to the dollar-amount wage subsidy case. They are thus less efficient in the short run than this last scheme because under them the government must pay for the increase in average labor costs as well as for the increase in employment.

If we modify some of our assumptions the dollar-amount wage subsidy is no longer uniformly most efficient. For example, even though the training in our model is assumed to be specific to the firm, it may have as a joint product the generally applicable ability of the worker to accept the discipline and routine of the workplace. This ability increases the worker's likelihood of finding another job when he leaves his present employment. If this ability increases with the duration of employment on the present job then the percentage wage subsidy, with its decreased quit rate and resulting greater average duration of employment, has some advantage. This benefit must, of course, be weighed against its greater cost in producing direct increases in employment. The training subsidy, on the other hand, results in a lower wage and a lessened duration of employment. Both on this ground and because of its greater cost it is inferior to the dollar-amount wage subsidy.

REFERENCES

P. Diamond, "Negative Taxes and the Poverty Problem," *Nat. Tax J.*, Sept. 1968, *21*, 288–303.

W. Oi, "Labor as a Quasi-Fixed Factor of Production," *J. Polit. Econ.*, Dec. 1962, *70*, 538–55.

Managerial Pay and Corporate Performance

By WILBUR G. LEWELLEN AND BLAINE HUNTSMAN*

At the core of most economic analyses of industrial behavior is the proposition that the managers of an enterprise guide its activities in such a way as to maximize the monetary well-being of its owners. The theory of the firm in its conventional form depends heavily on this presumption, and the alleged allocative efficiency of the private enterprise system is founded on its implications. Corporations can, of course, simultaneously pursue a variety of goals, with different top management groups placing differing degrees of importance on a range of alternative objectives. If, however, it should turn out that profit-related goals are consistently subordinated, the relevance of much of our received economic doctrine would become suspect.

That doctrine notwithstanding, the possibility has been raised in recent years that goals other than maximization of profits or share prices may well govern managerial behavior. The dissenting point of view can be traced to the descriptive arguments advanced by Adolph A. Berle and Gardiner C. Means nearly forty years ago. They asserted, in substance, that the rise of the giant corporation was increasingly serving to separate control from ownership. The natural corollary of this contention is that management is not necessarily constrained to act with the owners' welfare in mind, but can be expected instead to serve primarily its own economic self-interest. More recently, William J. Baumol (1958, 1962, 1967) has suggested that management can—and, in fact, does—cater to such self-interest by maximizing the total sales revenues of the enterprise, with profit increases being but a secondary issue. Baumol's argument is based at least in part, on the observation that executive salaries appear to be "... far more closely correlated with the scale of operations of the firm than with its profitability" (1967, p. 46). However, it is not clear a priori that the economic self-interests of ownership and management *can* be neatly separated. For example, most top executives have substantial stock holdings in their own firms and receive a sizeable fraction of their total compensation in the form of stock options and other stock-related pay arrangements (see Wilbur G. Lewellen 1968, 1969a).

The intent of the present paper is to examine empirically the question of whether the financial rewards reaped by high level corporate executives are more strongly influenced by company performance as measured by total corporate revenues or—alternatively—as measured by either of two standards of shareholder welfare. The results should provide a basis for drawing at least some tentative conclusions as to the expected nature of managerial behavior in the contemporary industrial environment.

We are aware that several previous empirical studies have been conducted with the same stated objective. Typically, the methodology has been to compare the respective degrees of correlation between executive compensation and firm sales on the one

* The authors are, respectively, associate professor of industrial management at Purdue University and associate professor of finance at the University of Utah. The research was supported with funds supplied by the National Bureau of Economic Research, the Purdue Research Foundation, and the Ford Foundation's grant for research in business finance to the Sloan School of Management at the Massachusetts Institute of Technology. The advice and counsel of Professors Frank Bass, George Horwich, Robert Johnson, and Edgar Pessemier of Purdue, Professor John Lintner of Harvard, and Professor Bruce Baird of Utah are gratefully acknowledged. The computations were performed at the computer centers of Purdue University and the Massachusetts Institute of Technology. Of course, the opinions and conclusions presented are solely the authors'.

hand, and executive compensation and firm profits on the other. Perhaps the most thorough of these studies is that by Joseph W. McGuire, John S. Y. Chiu, and Alvar O. Elbing[1] who concluded that sales and compensation were more highly correlated than were profits and compensation, and interpreted those results as supportive of the Baumol hypothesis. The findings presented in the current paper lead to interpretations that run counter to those of McGuire, et al., and are based on tests of a multivariate regression model which is designed to reduce the effects of various statistical and measurement biases. Our findings also reflect the use here of more comprehensive indices both of executive remuneration and shareholder welfare than have been characteristic of earlier investigations.

I. *The Sample*

The very large firm is obviously the one in which the phenomenon of the separation of ownership from control—and its possible consequences in terms of managerial behavior—is likely to be most severe. For that reason, the sample examined consists of 50 firms drawn from the top of the *Fortune* magazine list of the nation's 500 largest industrials. Virtually all the firms involved would be regarded as "blue chips" in the language of the investment community, all enjoy a very wide public distribution of their common stock, and a broad spectrum of industry categories is represented.

While data on such items as company sales and profits are readily accessible in published financial statements, certain other information which is essential to the proper execution of a study of the sort undertaken here is not so easily obtained. In particular, a truly comprehensive measure of managerial remuneration requires that the worth of *all* the major constituents of the relevant pay packages be recognized, including, in addition to cash salaries and bonuses, the full range of indirect, deferred, and contingent compensation arrangements that executives enjoy. The raw material for an evaluation of these devices for senior management is publicly available only in the proxy statements which corporations must submit in conjunction with their annual shareholders' meetings. Because of variations in the manner in which firms present these data, however, and because of changes in the SEC's reporting rules over the years, it is not always possible to put together an adequate record of top executive rewards for every large company one might care to inspect.[2] As a result, it became necessary to dip down as far as the corporation ranked 94th on *Fortune*'s most recent tabulation[3] in order to assemble a sample of 50 whose executives' earnings histories could be compiled accurately for any length of time. Those 50 firms currently account for approximately one-fourth of the total sales and profits of the entire manufacturing sector. They are listed in the Appendix.

The study examines the cross-sectional relationships between executive compensation and company performance at three-year intervals, beginning with 1942 and ending with 1963. This period was chosen because the requisite figures were available on total executive remuneration from an earlier investigation of the components of managerial pay by one of the authors (Lewellen 1968). Since similar information would have been exceedingly expensive to duplicate for other periods, as shown in Section III, its existence seemed a compelling argument in favor of exploiting that resource for our present purposes. The entire time span was analyzed in order to detect any inter-temporal evolution in the underlying relationships which might have arisen from phenomena such as continuing corporate growth. The years examined also encompass an era of significant structural change in the economy, especially in the tax environment, and include the pervasive influence of several business cycles and two wars.

[1] Others include those of Arch Patton, David R. Roberts, and Oliver E. Williamson.

[2] A discussion of the methods of compiling the data is contained in Section III.

[3] The 1968 list was the latest one available at the time the analysis here was initiated.

II. *The Model*

The major hypotheses which have been advanced to explain observed levels of senior executive compensation can be summarized in their simplest forms as follows: (i) A firm's top management is rewarded primarily on the basis of its ability to increase the profits of the enterprise; or (ii) The compensation of top executives depends upon the volume of sales generated by the corporation whose affairs they administer. Given the widely accepted proposition that the market value of a firm's equity at any point in time is approximately equal to the discounted worth of the future returns anticipated by actual and potential owners, (i) can be interpreted as the analogue of the share price maximization precept.

Of these hypotheses, (i) conforms with the behavioral assumptions underlying the conventional theory of the firm, while previous empirical findings (McGuire, Chiu, and Elbing, Patton, Roberts) appear to support (ii). Other possible viewpoints include the propositions that executives are rewarded according to *both* profit and sales performance, or, at the opposite extreme, that neither item is relevant. The model developed here is designed to test this range of alternatives.

The Basic Model

Abstracting, for the moment, from potential statistical and measurement problems, and in the absence of theoretical reasons to specify an alternative form of functional relationship, we may begin by postulating that a top executive's compensation is related in linear fashion to both the profits and sales of the firm he manages. The structural form of the relationship can be written:

$$(1) \qquad C_{it} = a_0 + a_1\pi_{it} + a_2S_{it} + u_{it}$$

Where, C, π, and S represent executive compensation, corporate profits, and corporate sales, respectively, and u is a random disturbance term. Subscript i denotes the firm and subscript t the period to which the measure applies.

By supplying a basis for observing the magnitude of the coefficients a_1 and a_2 and the levels of statistical significance attaching thereto, the above specification provides a natural vehicle for inferring the relative influence of the two independent variables upon compensation, and thereby testing the alternative hypotheses. The emergence of a positive value for the constant term a_0 would imply, in effect, that executive rewards rise less than in proportion to company sales and/or profits. Thus, it seems probable that a \$50 thousand difference in annual profits between two firms in the \$100 million profit range would result in a smaller difference in the pay of their respective chief executives than would the same dollar profit difference in the case of two firms whose yearly earnings were in the \$100 thousand range. Represented graphically, the compensation vs. profits or compensation vs. sales relationship would therefore be expected to be concave downward for a sample of enterprises differing widely in size, and the linear approximation to any segment of such an underlying relationship would necessarily include a positive intercept value. It follows, then, that a_1 and a_2 must be interpreted in marginal—although constant for the sample range—terms throughout.

Statistical Problems

Unfortunately, direct application of equation (1) to any generalized sample of cross-sectional data can be expected to encounter several possible sources of statistical bias. For one thing, the efficiency of least square estimates depends upon the variances of the disturbance terms being constant. Examination of scatter diagrams of pilot regression runs using equation (1) revealed, as anticipated, that the error terms were *not* constant but were approximately in proportion to the dependent variable. Moreover, and as one might also suspect, those firms relatively large by virtually any scale criterion were also characterized by relatively high sales and profits levels. This scale-associated linkage between the independent variables poses the threat of serious collinearity,[4] with re-

[4] The high degrees of correlation between the independent variables in their natural form were indicated by the presence of simple correlation coefficients which in most cases, exceeded .9. This high degree of observed

sultant difficulties in estimating the separate influences of those variables. The regression coefficients produced by fitting data to (1) were characterized by inordinately large standard errors, suggesting that the correlation between the independent variables was indeed too high to permit the generation of reliable estimates.

An approach which attacks both these problems is the weighted regression technique.[5] Since scatter diagrams of equation (1) indicated that the error terms (u_{it} for all i) tended to vary directly with the dependent variable, an appropriate weighting procedure is to divide each variable in (1) by any one of several scale-related deflators, provided there are acceptable grounds for expecting that the variances of the deflators bear a proportionate relationship to those of the u_{it}. Moreover, by creating ratios in which both numerator and denominator are associated with the firm's size, the weighted regression approach eliminates the basic reason for expecting high degrees of correlation between the variables as a consequence of a common scale factor. Following the Miller and Modigliani precedent, book value of total assets was the weighting factor chosen because the resulting deflated variables have "... natural and useful economic interpretations in their own right" (Merton H. Miller and Franco Modigliani 1966, p. 350).[6] Deflating (1) by total book assets, denoted by A, yields the following form:

$$(2)\qquad \frac{C_{it}}{A_{it}} = a_0\left(\frac{1}{A_{it}}\right) + a_1\left(\frac{\pi_{it}}{A_{it}}\right) + a_2\left(\frac{S_{it}}{A_{it}}\right) + u^*_{it}$$

Since the variances of the deflator selected are hypothesized to be roughly proportional to those of the u_{it}, the error terms in the original equation, we would expect the variances of the new error terms ($u^*_{it} = u_{it}/A_{it}$) to be approximately constant over the sample range. Examination of scatter diagrams indicated that the requisite least squares assumption of homoscedasticity was in fact met by (2). As to the collinearity problem, tests of the model evinced materially smaller standard errors of the coefficient estimates in addition to producing consistently high t-values in conjunction with tests of significance. Both these phenomena suggest that the deflated variables are not sufficiently collinear as to interfere with the formation of reliable coefficient estimates.[7]

III. *Measurement*

An empirical test of the model necessitates that meaningful measures of the separate variables be obtained. This section defines the measures chosen and discusses the rationale underlying their selection.

Executive Compensation

Previous studies seeking to determine the influence of corporate performance on executive compensation merit some skepticism in the sense that only *partial* indices of compensation, typically, cash salary plus bonus payments alone were employed (McGuire, Chiu, and Elbing, Patton, Roberts). Other forms of remuneration such as pension benefits, deferred pay, qualified profit-sharing plans, and stock options have been ignored entirely.

correlation is consistent with the results of a prior study by Bevars D. Mabry and David L. Siders, which employed correlation analysis to investigate the relationship between profits and sales.

[5] An alternate approach commonly used to ameliorate scale-related problems, given that there are no a priori grounds for specifying a linear vis-à-vis a multiplicative relationship, is to convert the natural values of the variables into logarithms. Results of the tests of the *log* transform of (1) were in fact characterized by somewhat lower simple correlation coefficients between $\log \pi_{it}$ and $\log S_{it}$, but the scatter diagrams still suggested the presence of excessive heteroscedasticity. The *log* transform was therefore rejected as a device for testing the hypothesis.

[6] That "natural and useful interpretation" in the present context relates to the fact that equation (2) now implicitly describes a process whereby management is viewed as maximizing company sales or profits per dollar of resources employed, i.e., maximizing subject to the available resource constraint.

[7] It might be noted that the model (2) is now linear homogeneous. Accordingly, consistency in interpretation requires that it be tested with the constant term suppressed. In order to check this specification, we first ran the regression with a constant term added and found it in no case to be significantly different from zero.

These latter items, however, frequently result in annual after-tax increments to executives' wealth that are of magnitudes many times the concurrent salary and bonus awards.

The proper treatment of deferred and contingent elements represents a task of major proportions. Lewellen's 1968 study embodies the relevant analysis and provides comprehensive measures of total executive earnings within the sample companies. The approach employed in appraising the worth of the noncash components of managerial remuneration was to calculate, for each individual senior executive and for each year in which he appeared in his firm's proxy statements, what might be termed the "current income equivalent" of his various deferred and contingent pay schemes. This consisted of the amount of additional direct cash income he would have required from his company to be as well rewarded after taxes as he was by all the supplemental compensation arrangements he actually enjoyed. In the case of a pension plan, for example, the procedure was to determine the additional salary or bonus the executive would have needed if he were to be enabled, with those additions, to purchase for himself an individual retirement annuity from an insurance company similar in form and equal in value to the benefits promised him under his employer's retirement plan. Similar calculations were made for other fringe benefits and the aggregate worth of his entire compensation package established. The total remuneration figures used in the present study therefore represent the sum of all such estimated current income equivalents plus salary and bonus.[8]

As a matter of both convenience and efficiency, the compensation of the single highest-paid executive in each firm for each year is taken into account here as the dependent variable measuring the size of the firm's managerial pay package. While it may seem more appropriate that the remuneration of *all* the senior policy-making individuals in a corporation be tested for a relationship to company performance, it happens that the pay of a firm's top man is a suitable surrogate for the pay of his closest subordinates in terms of their relative standing vis-à-vis corresponding officials in other firms.[9]

For comparability with prior empirical work, the model was initially tested using only the salary and bonus receipts of the senior officer of each company as the dependent variable, and then tested again using the more comprehensive total compensation measure instead. The symbol C_{i_t} is employed below to denote the salary plus bonus, and $C^*_{i_t}$ to denote the total remuneration, of the top executive of firm i in year t.

Measurement of Profitability

The model was also subjected to two different sets of tests, each using a separate index of shareholder welfare as an independent variable. The first set incorporated a direct measure of book profits while the second employed an equity market value measure. The book profit choice is perhaps the more conventional, and provides a basis for comparing our findings with those of earlier writers. Nevertheless, some caution must be exercised in interpreting regression results which depend on such a measure.

There are statistical drawbacks associated with the use of reported profit which stem from both its determination and behavior as contrasted with the same features of the other hypothesized performance criterion—namely, corporate sales revenues. While relatively uniform bases exist for recording revenues,[10] profit measurement is conditioned by a range of accounting options which

[8] The relevant computations were made by matching throughout the after-tax present value of the actual compensation arrangement being considered on the one hand, and its contrived current income equivalent on the other. For the details, see Lewellen (1968, ch. 2–6). A similar computational approach can be found in a study by Leonard R. Burgess.

[9] When the sample corporations were ranked in selected years first according to the total compensation of their top executive alone, then the total for their top three executives combined, and finally the total for their top five executives combined, the Spearman rank correlation coefficients between the three schedules were consistently on the order of .95 and were significant at the .0001 level in all instances (see Lewellen 1968, pp. 227–28).

[10] The definition employed here is the standard one of gross sales, less discounts, returns, and allowances.

are much less uniform. The areas of depreciation policy and inventory valuation offer prominent examples. Additionally, an examination of the year-to-year changes in reported corporate profits vis-à-vis sales changes indicates that the former measure is much more sensitive to short-run economic influences than is the latter, and consequently is more volatile. Profits therefore are more apt to depart from their "true" or "normal" values when observed in any given year than are sales, giving rise to the likelihood that the profits coefficients in regression equations of the form being tested here will tend to be biased downward.[11] But, because in our tests profits proved consistently to be a more powerful explanatory variable than did sales, steps either to adjust the profit measure for accounting disparities among firms or to remove some of the random noise by various normalization procedures were not taken. If they had been, the only likely effect would be to increase further the size and, presumably, the significance of the profits coefficients. Such an effect would in no way alter the interpretation of our results.

Besides these statistical problems, the use of reported profits for any particular year as the index of corporate achievement implies that no conflict exists between short- and long-run maximization strategies by management, i.e., that there is no real trade-off between increasing the current year's net income and increasing the current value of future income. There are, however, a variety of circumstances which could lead to such a conflict. For example, a cutback in expenditures for the proper maintenance of plant and equipment can produce very favorable current profit results, but at the sacrifice of subsequent earnings because of the deterioration of physical assets. Conversely, expensed outlays for research into new product opportunities will reduce immediate reported profits but can provide the foundation for sizeable future earnings. Growth-oriented decisions, depending on the market expectations they generate, may furnish substantive current benefits to shareholders by increasing the market value of their holdings. The recent literature of business finance is, of course, rich in its emphasis on the maximization of share prices as the correct managerial goal (see David Durand, Myron J. Gordon, Lewellen (1969 b), John Lintner, Modigliani and Miller (1958), Alexander A. Robichek and Stewart C. Myers). It was for this reason that the retests using equity market value (denoted below by V_{i_t}) in place of profits were conducted.

[11] A well-publicized single-equation model which is characterized by two independent variables, one subject to both measurement errors and random fluctuations, is that commonly used to test the relative influence of dividends and retained earnings on share prices. For a detailed discussion of the procedures for dealing with these phenomena in the context of the dividend policy equation, see Irwin Friend and Marshall Puckett.

Measurement of Assets

As will be recalled, a measure of total book assets was introduced into the final version of the model (2) as a deflator. The asset measure used, denoted A_{i_t}, is defined net of depreciation for firm i as of the beginning of year t.

Summary of the Variable Definitions

The notation adopted, then, is as follows:

C_{i_t} = Salary plus bonus payments received by the chief executive of firm i during year t;

$C^*_{i_t}$ = Total after-tax compensation of that executive, including the equivalent value of all major deferred and contingent compensation arrangements;

π_{i_t} = Reported total after-tax profits of firm i during year t;

S_{i_t} = Total sales revenues of the firm, net of discounts, returns, and allowances;

V_{i_t} = Total market value of the firm's outstanding common stock as of the beginning of year t;

A_{i_t} = Total book value of the assets of the firm, net of depreciation, at the beginning of year t.

These measures, as formulated from the data gathered from the 50 corporations in the sample, were fitted to the model developed in the preceding section.

IV. *The Empirical Results*

Table 1 shows the results of the initial tests which were conducted using profits and sales as the measures of corporate performance. Table 2 displays the results of the retests in which equity market value was substituted for profits. Both tables are divided into two panels. The left-hand panel records the findings when salary plus bonus payments alone are adopted as the measure of chief executive compensation, while the right-hand side indicates the results when total executive pay is the dependent variable.

Profits, Sales, and Compensation

The evidence in Table 1 provides strong support for the hypothesis that top management's remuneration is heavily dependent upon the generation of profits. In particular, the signs of the coefficients of the profit measure are positive for each cross-section regardless of the compensation measure employed. Moreover, the profit variable proves highly significant for all runs in which the dependent variable consists of executive salary plus bonus payments, and for six of the eight years in the case of total compensation as the dependent variable. By contrast, the sales measure is in no instance statistically significant, and the sales coefficients are approximately equally divided among years with respect to sign.

In view of the previously cited findings of earlier studies (see McGuire, Chiu, and Elbing, Roberts) in which the compensation measure adopted was essentially the same as our C_{i_t} (top executive salary plus bonus), we were more than mildly surprised at the apparent strong influence of π_{i_t} on C_{i_t} here, as well as at the accompanying lack of any statistically significant relationship between C_{i_t} and S_{i_t} (the left-hand panel of Table 1).[12]

[12] We were also surprised—especially in the light of the scaled nature of our model—at the rather high degree of explanatory power it displayed. The coefficients of multiple correlation exceeded .9 for the five most recent years tested, and were nowhere lower than .737.

TABLE 1—REGRESSION RESULTS: PROFITS, SALES, AND COMPENSATION

Year	Regression Equation: $\frac{C_{i_t}}{A_i} = a_0\left(\frac{1}{A_{i_t}}\right) + a_1\left(\frac{\pi_{i_t}}{A_{i_t}}\right) + a_2\left(\frac{S_{i_t}}{A_{i_t}}\right) + u^*_{i_t}$							Regression Equation: $\frac{C^*_{i_t}}{A_{i_t}} = a_0\left(\frac{1}{A_{i_t}}\right) + a_1\left(\frac{\pi_{i_t}}{A_{i_t}}\right) + a_2\left(\frac{S_{i_t}}{A_{i_t}}\right) + u^*_{i_t}$						
	a_0	t	a_1	t	a_2	t	R^2	a_0	t	a_1	t	a_2	t	R^2
1942	95.3[a]	8.86	5039.5	2.35	−86.6	1.20	.806	40.1[a]	13.99	1928.1	3.37	−28.6	1.49	.912
1945	101.0	6.52	3307.2	1.98	−50.8	.81	.744	41.9	7.52	1610.8	2.68	−27.6	1.21	.798
1948	86.7	8.38	1413.9	2.48	12.6	.30	.833	73.0	9.63	447.7	1.07	6.6	.20	.830
1951	109.9	12.89	1397.0	3.04	− 5.6	.20	.927	68.1	9.58	1202.9	3.14	−14.6	.65	.880
1954	129.3	9.68	1192.5	3.27	−23.1	.97	.910	85.8	4.55	1188.7	2.13	− 8.7	.26	.751
1957	112.9	7.59	963.5	3.89	6.6	.40	.929	146.6	1.96	461.6	.37	30.6	.37	.439
1960	121.3	6.65	660.7	3.12	15.1	.88	.919	130.5	3.40	1181.2	2.65	1.8	.00	.761
1963	155.3	9.24	677.5	3.91	−15.6	1.08	.932	71.8	2.35	875.8	2.77	17.4	.66	.737

[a] Times 10^3. For purposes of the regression runs, executive compensation was measured in actual dollars and all other variables were measured in millions of dollars. The economic interpretation of the profit coefficient (a_1) listed for 1963—to take an example—therefore is as follows: For corporations in the size range encompassed by the sample, every \$1 million differential in reported company profit was, on the average, accompanied by a \$677.50 differential in the annual pre-tax salary and bonus earnings of the firm's chief executive.

TABLE 2—REGRESSION RESULTS: MARKET VALUES, SALES, AND COMPENSATION

Year	Regression Equation: $\frac{C_{i_t}}{A_{i_t}} = a_0\left(\frac{1}{A_{i_t}}\right) + a_1\left(\frac{V_{i_t}}{A_{i_t}}\right) + a_2\left(\frac{S_{i_t}}{A_{i_t}}\right) + u^*_{i_t}$							Regression Equation: $\frac{C^*_{i_t}}{A_{i_t}} = a_0\left(\frac{1}{A_{i_t}}\right) + a_1\left(\frac{V_{i_t}}{A_{i_t}}\right) + a_2\left(\frac{S_{i_t}}{A_{i_t}}\right) + u^*_{i_t}$						
	a_0	t	a_1	t	a_2	t	R^2	a_0	t	a_1	t	a_2	t	R^2
1942	98.0[a]	9.72	339.2	2.82	−25.3	.48	.815	41.5[a]	15.08	109.9	3.35	− 1.5	.10	.912
1945	98.9	6.63	263.9	2.78	−33.5	.63	.762	41.8	7.61	103.3	2.95	−13.6	.70	.803
1948	86.9	8.75	239.9	3.25	9.0	.24	.846	73.3	9.66	51.9	.92	12.9	.45	.829
1951	111.1	12.81	105.3	2.73	14.1	.58	.925	69.1	9.35	78.1	2.37	6.8	.33	.870
1954	128.6	9.78	101.0	3.55	− 4.3	.22	.913	85.2	4.49	88.6	2.16	13.6	.47	.748
1957	118.8	7.63	39.8	3.13	19.4	1.20	.922	148.7	2.01	26.6	.44	33.5	.44	.440
1960	126.7	6.99	23.5	3.07	17.7	1.05	.919	138.1	3.93	59.7	4.03	− 4.4	.14	.796
1963	160.1	9.68	28.9	4.12	−10.1	.74	.934	78.3	2.72	45.6	3.73	19.1	.81	.766

[a] Times 10^3. Again for these runs, compensation was measured in dollars, and all other variables in millions of dollars. As indicated in connection with Table 1, the economic interpretation of the market value coefficient for, say, 1963 in the left-hand panel would take the following form: For companies in the relevant size range, a differential of \$1 million in the firm's total equity market value was accompanied by a \$28.90 differential in the annual pre-tax salary plus bonus enjoyed by its top executive.

Those earlier studies had led us to anticipate an initially weak relationship between profits and remuneration—or perhaps none at all—and, in addition, to expect that corporate sales would show up as the variable more strongly influencing managerial rewards. In other words, we were prepared to find that our first attempts would at best provide a basis for comparing the results of subsequent runs using more highly developed measures of the relevant variables, and that any major insights would stem from the introduction of such improved measures.

It is not likely that differences in either the sample bases tested or the time periods examined can account for the substantive differences between the findings presented here and those of Roberts and of McGuire, Chiu, and Elbing. All three samples were drawn from the nation's largest industrial firms, and the time span covered in the current study encompasses the years tested in both of the previous investigations. A more plausible explanation lies in the fact that the tests here were conducted within the framework of a more completely developed multivariate model designed to cope with the serious statistical problems discussed in Section II.

A second unexpected feature of the test results was the lack of improvement in fit upon substitution of total executive compensation for salary plus bonus as the dependent variable (right-hand panel of Table 1). Rather, both the levels of statistical significance associated with profits and the coefficients of multiple correlation *diminished* slightly for most years when C_{i_t} was replaced by $C^*_{i_t}$. There are at least two possible explanations for this phenomenon: (i) The performance of the model using the partial compensation measure, C_{i_t}, was substantially better than anticipated, leaving relatively little room for improvement; (ii) A sizeable proportion of $C^*_{i_t}$, especially in the later years tested, is attributable to compensation arrangements whose values depend importantly on short-term fluctuations in employer-company share prices—stock options, for example. Because these short-term fluctuations reflect a host of random influences,

the total compensation figures embody a considerable "noise" component.[13]

The general pattern of changes in both sets of results over time highlights an additional point. Looking again at the left-hand panel of Table 1, it can be seen that the *size* of the profit coefficients shows a marked year-to-year decrease for seven of the eight periods tested (a_1 shows a very slight increase between 1960 and 1963). Similarly, in the right-hand panel, excluding the only two runs for which the profit variable is not statistically significant (1948 and 1957), a corresponding secular decline is evident. Since the companies in the sample have grown over time, the simultaneously decreasing size of the profit coefficients supports our earlier expectation that the compensation vs. profits relationship is concave downward. The increasing values of a_0 displayed in the Table reflect the same phenomenon.

To summarize: the results shown in Table 1 indicate that equations utilizing salary plus bonus payments alone as a measure of executive remuneration yield slightly better regression fits than do those employing total compensation. But regardless of which compensation measure is adopted, reported company profits appear to have a strong and persistent influence on executive rewards, whereas sales seem to have little, if any, such impact.

Market Values, Sales, and Compensation

Because of the possible conceptual and measurement difficulties that are associated with annual book profit as a direct index of shareholder welfare, the model was retested using the alternative measure discussed in Section III, the total market value of the firm's outstanding common stock. This measure was chosen because it presumably incorporates the investing public's evaluation of future as well as current returns to owners, and also avoids the potential statistical problems that arise from differences in accounting procedures among firms.

The results of the retests are set forth in Table 2 and parallel those of the initial tests in virtually all relevant respects, with market value now appearing to be a major factor in the determination of compensation levels. Specifically, the coefficients of the market value variable have positive signs for all cross-sections no matter which measure of compensation serves as the dependent variable. In addition, the market value measure proves highly significant for all eight years when salary plus bonus is the dependent variable, and for six of the eight years when the performance measures are regressed against total compensation. Again, the sales variable is statistically insignificant in all cases and the sales coefficients vary in sign.

The total explanatory power of the equations containing equity market value as a control variable roughly matched that of the corresponding equations in which book profit was employed. The historical pattern of the retest results is also similar to that of the initial tests in that the secularly declining size of the market value coefficients and the secularly increasing a_0 values suggest downward concavity in the underlying relationship. Thus, the substitution of equity market value for profits in the equations has no material impact on either the nature or the interpretation of the findings.

V. *Summary and Conclusions*

The question of whether a corporation's profitability or its sales revenue has the stronger influence on the rewards of its senior officers has been examined here by means of a multivariate analysis. The underlying issue is that of the personal payoff to the professional manager for pursuing operating objectives that enhance the monetary well-being of shareholders. Because the results of the study persistently indicate that both reported profits and equity market

[13] In the early years shown, the bulk of senior executive compensation consisted of salary and bonus payments. Consequently, for these early years the results of the tests empolying C_{it}^* do not differ materially from those in which C_{it} is the dependent variable. However, as the disparity between the partial and total compensation measures widens over time, i.e., as rewards other than salary and bonus become progressively more important in the total pay package, the regression results also diverge. For a discussion of the indicated shift in emphasis within the managerial compensation package, see Lewellen (1968).

values are substantially more important in the determination of executive compensation than are sales—indeed, sales seem to be quite irrelevant—the clear inference is that there is a greater incentive for management to shape its decision rules in a manner consonant with shareholder interests than to seek the alternative goal of revenue maximization. The evidence presented therefore can be interpreted as support for the notion that a highly-industrialized economy characterized by a diverse set of suppliers of capital, sizeable aggregations of productive resources, and a professional managerial class can in large measure still be analyzed by models which are based on the assumption of profit-seeking behavior.

Appendix

Companies in the Sample

Allied Chemical
American Can
American Cyanamid
American Metal Climax
American Tobacco
Anaconda
Bendix
Bethlehem Steel
Boeing
Borden
Caterpillar Tractor
Cities Service
Continental Can
Continental Oil
Douglas Aircraft
Dow Chemical
DuPont
Eastman Kodak
Firestone Tire
General Electric
General Foods
General Motors
General Tire
B.F. Goodrich
Goodyear Tire
Gulf Oil
Inland Steel
IBM
International Harvester
International Paper
IT&T
Jones and Laughlin Steel
Lockheed Aircraft
National Dairy Products
North American Aviation
Phillips Petroleum
Procter & Gamble
RCA
Republic Steel
Reynolds Tobacco
Shell Oil
Sinclair Oil
Standard Oil (Indiana)
Swift
Texaco
Tidewater Oil
United Aircraft
U.S. Rubber
U.S. Steel
Westinghouse Electric

References

W. J. Baumol, "On the Theory of Oligopoly," *Economica*, Aug. 1958, *25*, 187–98.

———, "On the Theory of Expansion of the Firm," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1078–87.

———, *Business Behavior, Value, and Growth*, rev. ed., New York 1967.

A. A. Berle and G. C. Means, *The Modern Corporation and Private Property*, New York 1934.

L. R. Burgess, *Top Executive Pay Package*, New York 1963.

D. Durand, "Costs of Debt and Equity Funds for Business: Trends and Problems of Measurement," *Conference on Research in Business Finance*, New York 1952, 215–47.

I. Friend and M. Puckett, "Dividends and Stock Prices," *Amer. Econ. Rev.*, Sept. 1964, *54*, 656–82.

M. J. Gordon, *The Investment, Financing, and Valuation of the Corporation*, Homewood 1962.

W. G. Lewellen, *Executive Compensation in Large Industrial Corporations*, New York 1968.

———, 1969a, *The Cost of Capital*, Belmont, Calif. 1969.

———, 1969b, "Management and Ownership in the Large Firm," *J. Finance*, May 1969, *24*, 299–322.

J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.*, Feb. 1965, *47*, 13–37.

B. D. Mabry and D. L. Siders, "An Empirical Test of the Sales Maximization Hypothesis," *Southern Econ. J.*, Jan. 1967, *33*, 367–77.

J. W. McGuire, J. S. Y. Chiu, and A. O. Elbing, "Executive Incomes, Sales, and Profits," *Amer. Econ. Rev.*, Sept. 1962, *52*, 753–61.

M. H. Miller and F. Modigliani, "Some Estimates of the Cost of Capital to the Electric Utility Industry," *Amer. Econ. Rev.*, June 1966, *56*, 333–91.

F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

A. Patton, *Men, Money, and Motivation*, New York 1961.

D. R. Roberts, *Executive Compensation*, Glencoe, Ill., 1959.

A. A. Robichek and S. C. Myers, *Optimal Financing Decisions*, Englewood Cliffs 1965.

O. E. Williamson, *The Economics of Discretionary Behavior*, Englewood Cliffs 1964.

Fortune, "The Fortune Directory of the 500 Largest U.S. Industrial Corporations," June 1968, *77*, 186–204.

Devaluation Risk and Forward Exchange Theory

By DON SCHILLING*

There is considerable current interest in policy problems which involve the foreign exchange markets. An important element in these problems is the role of speculation in forward exchange. The behavior of forward speculation is dependent upon the expected value of speculators' expectations of the future spot rate and upon the stochastic structure of these expectations. However, the structural aspect of speculative expectations has been largely neglected in the theoretical analysis of forward exchange.

Existing formal models of the forward exchange market have either assumed that speculators utilize only a point estimate of the expected spot rate or have assumed that they consider only the expected value and variance of some subjective distribution which is either left unspecified or assumed to be normal.[1] Although the normal distribution is not necessarily an inappropriate characterization of speculative expectations within a flexible rate market system, its applicability to the existing system of a narrow band centered around a fixed but adjustable peg is dubious. It has been generally recognized that speculative expectations under the present system are essentially binomial; either there will be a relatively large devaluation or there will be no change in the peg.[2] This paper presents a formal model of forward exchange market speculation which is based upon the present structure of the international monetary system with binomial speculative expectations.

* Assistant professor of economics at the University of Missouri. I would like to thank H. Latane for his assistance in the formulation of the model; and S. R. Johnson and an anonymous referee of this *Review* for their helpful comments while reserving for myself full responsibility for any errors.

[1] S. Black, P. B. Kenen, and H. R. Stoll use a point estimate; H. G. Grubel makes use of the mean and variance of an unspecified distribution; S. C. Tsiang and M. S. Feldstein assume that speculative expectations are normally distributed.

[2] L. B. Yeager (p. 207); J. Black; Feldstein (p. 184n).

Consider a simple form of the interest rate parity theory (*IRPT*) model in which there are only two currencies, the dollar ($) and the pound (£), which are traded against each other spot for immediate delivery and forward for delivery in ninety days. At any point in time there exist two $-£ exchange rates, the spot rate and the (90 day) forward rate.

In order to develop the relationship between the spot and forward rates, let us consider the interaction of two groups of market participants: 1) interest rate arbitragers who seek to profit from international yield differentials on short-term liquid assets without exposing themselves to the risk of adverse exchange rate changes, by offsetting or "hedging" every spot purchase of foreign exchange with an equal sale of forward exchange; and 2) forward speculators who undertake speculative commitments in forward exchange in hope of profiting from the differential between the present forward rate and their expectation of what the spot rate will be when their forward contract matures. Because interest rate arbitragers sell forward the higher yielding currency, they tend to drive the forward rate on that currency below its spot rate. If there were no forward speculators in the market, the forward rate would be driven to its interest rate parity where it is just sufficiently below the spot rate to exactly offset the interest rate differential.[3] If however, speculators have little or no doubt of the intent and ability of the higher yielding currency's exchange authority to maintain its par value in the immediate future, they will buy the dis-

[3] The interest parity forward rate may be computed as a percent of the spot rate by the following formula,

$$x_p/x_0 = (1 + i_{\$}t)/(1 + i_{£}t)$$

where x_p is the parity rate, x_0 the spot rate, $i_{\$}$ and $i_{£}$ are, respectively, the *U.S.* and the *U.K.* interest rates, and t is the maturity period of the forward contract as a fraction of a year.

counted forward exchange as it drops below the spot rate. These activities will normally balance each other at a point somewhere between the spot and parity forward rates; its exact position is dependent upon the relative elasticities of speculation and arbitrage.

II. *The Speculative Demand Function*

Within the present foreign exchange system nearly all spot exchange rates are stabilized to within $\pm\frac{3}{4}$ percent of their par value in terms of the dollar. Therefore, unless there is a change in the par value, the maximum loss to a speculator is a very small percentage of his position. If there is a change in par value, it will be very likely to take the form of a devaluation of one currency of 15 percent or more. For developed countries, devaluation has become the policy measure of last resort. A country's balance of payments is usually in serious and chronic deficit by the time devaluation is imposed. A large devaluation is therefore needed in order first to achieve a current account surplus with which to rebuild depleted currency reserves and second, to convince speculators that further devaluation will not be necessary in the immediate future.

The immediate reaction of a central bank when confronted with a balance of payments deficit is to raise short-term domestic interest rates in order to discourage domestically owned funds from leaving and to encourage an inflow of foreign interest arbitrage funds. The monetary authorities and government of the affected country would normally introduce a sequence of increasingly stringent and more direct policy measures aimed at stemming the deficit before reaching the last resort. Thus the period between raising the interest rate to a level sufficiently high to induce a forward parity discount and devaluation will ordinarily be at least ninety days.[4]

These characteristics of present day foreign exchange markets lend support to the proposition that speculative risk is essentially binomial. Either the spot rate on the higher yielding currency in ninety days will be quite close to its present level, or it will be greatly devalued in the direction of the interest rate parity discounted forward rate. In turn, this suggests a basic assumption of our model, that speculative expectations can be approximated with a binomial subjective probability distribution.

It is further assumed that the forward exchange speculator is a risk-averse expected utility maximizer who has the specific utility function $U(\text{wealth}) = ln$ (wealth). The use of this specific utility function is justified primarily on grounds of analytical convenience in allowing an explicit analytical solution for an individual speculator's excess demand curve for forward exchange. In addition it provides an efficient speculative strategy in that consistent maximization of expected utility implies a growth path of wealth which asymptotically approaches maximization of the compound average annual average return on wealth as the number of decision periods becomes very large.[5] Also, the *ln* utility function possesses the useful characteristic that it has a unitary coefficient of relative risk aversion, making the optimum (expected utility maximizing) solution independent of the level of the initial wealth of the speculator and allowing explicit aggregation from individual to market functions.[6]

In order to simplify the analysis, it is assumed that the flow demands of importers and exporters are completely insensitive to the spot and forward exchange rates so that variations and imbalances in these quantities

[4] Available information was examined for the 1931 devaluation of the pound, the 1934 devaluation of the dollar, 1949 devaluation of the pound, and the 1967 devaluation of the pound. In all, twenty-eight markets were examined ninety days before devaluation and in only two cases (the Italian lire-sterling market in 1931, and the Canadian dollar—sterling market in 1967) was the devalued currency at a forward premium 90 days before its devaluation. During the last forty years the only upward revaluations of a major currency have been the 1961 and 1969 revaluations of the Deutschemark. Ninety days preceding each of its revaluations the mark was at a premium in all markets except for the *U.S.* dollar market in 1961 where it was at a very slight discount.

The data were obtained from P. Einzig pp. 448–81 (1931 devaluation) and from appropriate issues of *The (London) Times* (1949 devaluation), *Journal of Commerce* (1934 devaluation) and *Montigue's Review* (1967 devaluation).

[5] See Henry Latane.

[6] The coefficient of relative risk aversion is defined as

become purely exogenous influences upon the financial equilibrium developed in the model. Although this assumption is quite restrictive, it is not an unreasonable approximation of reality. These demands are not likely to be very responsive to the type of variations in the spot and forward rates that are possible when the spot rate cannot move outside of a 1.5 percent range.[7]

The following assumptions are made for analytical and expositional convenience: 1) Speculators are assumed to believe (with certainty) that the spot rate in ninety days will be equal to the present spot rate if devaluation does not occur and they have a specific (large) amount of devaluation in mind if devaluation does occur;[8] 2) Interest rate arbitrage transactions are taken to consist of simultaneously purchasing one exchange spot, selling it forward, and investing the spot exchange in short-term government bills for the life of the forward contract; 3) It is assumed that no deposit is required for the purchase of a forward contract and transactions costs are ignored; 4) Speculative transactions are treated solely in terms of the outright purchase or sale of forward currency.[9]

Consider a situation in which the treasury bill rate in the United Kingdom is greater than in the United States. In response to the activities of interest rate arbitragers, sterling is at a discount with respect to the dollar. The optimal proportion of a speculator's wealth to be committed to a forward exchange position may be stated as a function of his subjective evaluation of devaluation risk and his expected yield on position. To derive this result define:

$$(1)\quad a = \frac{x_3 - x_1}{x_1} = \text{the expected gain per dollar of commitment in forward sterling if devaluation does not occur.}$$

$$(2)\quad b = \frac{x_2 - x_1}{x_1} = \text{the expected gain per dollar of commitment in forward sterling if devaluation does occur.}$$

where x_1 is the present price of forward pounds, x_2 is the expected price of spot sterling in 90 days if devaluation does occur, and x_3 is the expected price of spot sterling if devaluation does not occur. x_3 is initially assumed to be equal to x_0, the present spot rate.

Let y equal the speculator's commitment in forward exchange as a proportion of w, his total wealth.[10] If devaluation does not occur, the speculator's wealth will become $w(1+ya)$,

$-[U''(x)/U'(x)]x$, which is equal to negative unity if $U(x) = ln\ x$.

[7] A trader can alter his flow demand through the foreign exchange market only by altering the overall level of his business activity or by shifting his business between domestic and foreign operations. Therefore, changing the level of a flow is likely to be slow and costly relative to the stock changes covered in the model. As a result, they are unlikely to be undertaken in response to the small and transient changes to which the spot and forward exchange rates are limited under the present system.

Note that stock demands of traders resulting from changes in inventories and holdings of short-term financial claims can be (and probably are) highly responsive to small changes in the exchange rates. However, these are equivalent in their market effect to the operations of speculators and arbitragers.

[8] The assumption of expectational certainty about each of these binomial values is obviously not literally true. It is intended to be a convenient and reasonably efficient approximation of a very bimodal distribution whose density is sharply concentrated around one mode within the present intervention points and less densely dispersed around some expectation of the new par if devaluation should occur.

[9] Because the forward market effects of stock adjustments by traders, uncovered arbitragers, commercial banks and indeed all market participants other than pure arbitragers and pure speculators are ignored in the structurally simple model presented here, the analysis would appear to be lacking in generality. However, I have shown in my *J. Finance* article that it is possible to generalize this simple structure to include these actions in such a way that the diagrammatic analysis remains essentially unchanged. Only the quantity dimension of the analysis is affected and it is redefined in such a way that a comparative static shift results in a flow of funds defined in the broadest sense rather than merely a flow of covered arbitrage. Because our primary concern is with the structure of speculative expectations rather than the structure of forward exchange market models, the present approach was selected for analytical simplicity.

[10] Under the present assumption of no speculative deposit requirement, y is unconstrained. If a deposit were required its range of variation would be $-w/d < y \leq w/d$ where w is the speculator's wealth and d is the deposit requirement.

and if it does occur, his wealth will become $w(1+yb)$. Let P be the speculator's subjective probability of a devaluation occurring during the ninety day period of his forward contract. The expected utility of a forward sterling position equal to y of the speculator's wealth is:

$$(3)\quad E(U)=(1-P)\ ln\ (1+ya)+P\ ln\ (1+yb)$$

For given values of a, b, and P, the first derivative of (3) with respect to y is

$$(4)\qquad \frac{dE(U)}{dy}=\frac{(1-P)a}{1+ya}+\frac{Pb}{1+yb}$$

Now, set $dE(U)/dy=0$ and solve for the expected utility maximizing y designated as y^*.

$$(5)\qquad y^*=\frac{P}{b}-\frac{P}{a}-\frac{1}{b}$$

Substituting equations (1) and (2) into (5) and simplifying gives

$$(6)\qquad y^*=x_1\left[\frac{P-1}{x_2-x_1}-\frac{P}{x_3-x_1}\right]$$

Equation (6) may be interpreted as representing the speculator's optimum forward commitment as a proportion of his total wealth. It is stated as a function of the present forward rate for given levels of x_2, x_3, and P.

The speculator's indifference forward rate ($\bar{x}_1$) at which he will neither buy nor sell forward sterling may be obtained from equation (6) by setting y^* equal to zero and solving for x_1.

$$(7)\qquad \bar{x}_1=x_3+P(x_2-x_3)$$

For any given level of P, x_2, and x_3:

If $x_1<\bar{x}_1$ then y^* will be greater than zero. The optimum commitment will involve the purchase of forward sterling.

If $x_1>\bar{x}_1$ then y^* will be less than zero. The optimum commitment will involve the sale of forward sterling.

The aggregation of the individual speculators' demand functions to produce a market speculative excess demand curve requires some additional assumptions.

Assume that there are n speculators in the forward exchange market. Each speculator has the same subjective estimate of x_2 and x_3, but they may have different levels of wealth and differing subjective evaluations of the probability of devaluation. Let each speculator's wealth be expressed in terms of some common measure of wealth with each speculator maximizing $U(x)$. The speculators' aggregate excess demand function can then be expressed as

$$(8)\quad Y^*=\sum_{i=1}^{n} y_i^* w_i=x_1\left[\frac{\bar{P}-1}{x_2-x_1}-\frac{\bar{P}}{x_3-x_1}\right]K$$

Where Y^* is the aggregate excess demand of the n speculators for forward exchange, K is their aggregate wealth, and $\bar{P}$ is the wealth-weighted mean of their subjective probabilities of devaluation.

III. *The Model*

A graphic model of the forward exchange market is presented in Figure 1. Forward and spot exchange rates are plotted on the vertical axis. The horizontal axis represents the outstanding quantity of forward contracts between speculators and arbitragers and, because of the quantity equivalence between the spot and forward transactions of pure interest rate arbitragers, it also represents the stock of arbitrage funds held abroad under forward cover.

The two speculative curves, SS and $S'S'$, are mappings of equation (8) for $\bar{P}$ equal to .001 and .1, respectively, and are presented in the form of excess demand curves with x_3 equated to the present spot rate (x_0), and x_2 assumed equal to $.77x_0$.[11]

The interest rate arbitragers' curve (AA) appears as an excess supply curve of forward exchange. Its shape is justified by simply noting that large internationally oriented banks and large internationally oriented corporations at any given time have very considerable amounts of liquid funds at their disposal which they are willing to arbitrage under forward cover for a marginally greater yield. To a bank or corporation, however,

[11] The exact value of x_2 is arbitrary. Twenty-three percent happened to be the unweighted mean of eight devaluations of developed countries' currencies that took place between 1945 and 1965.

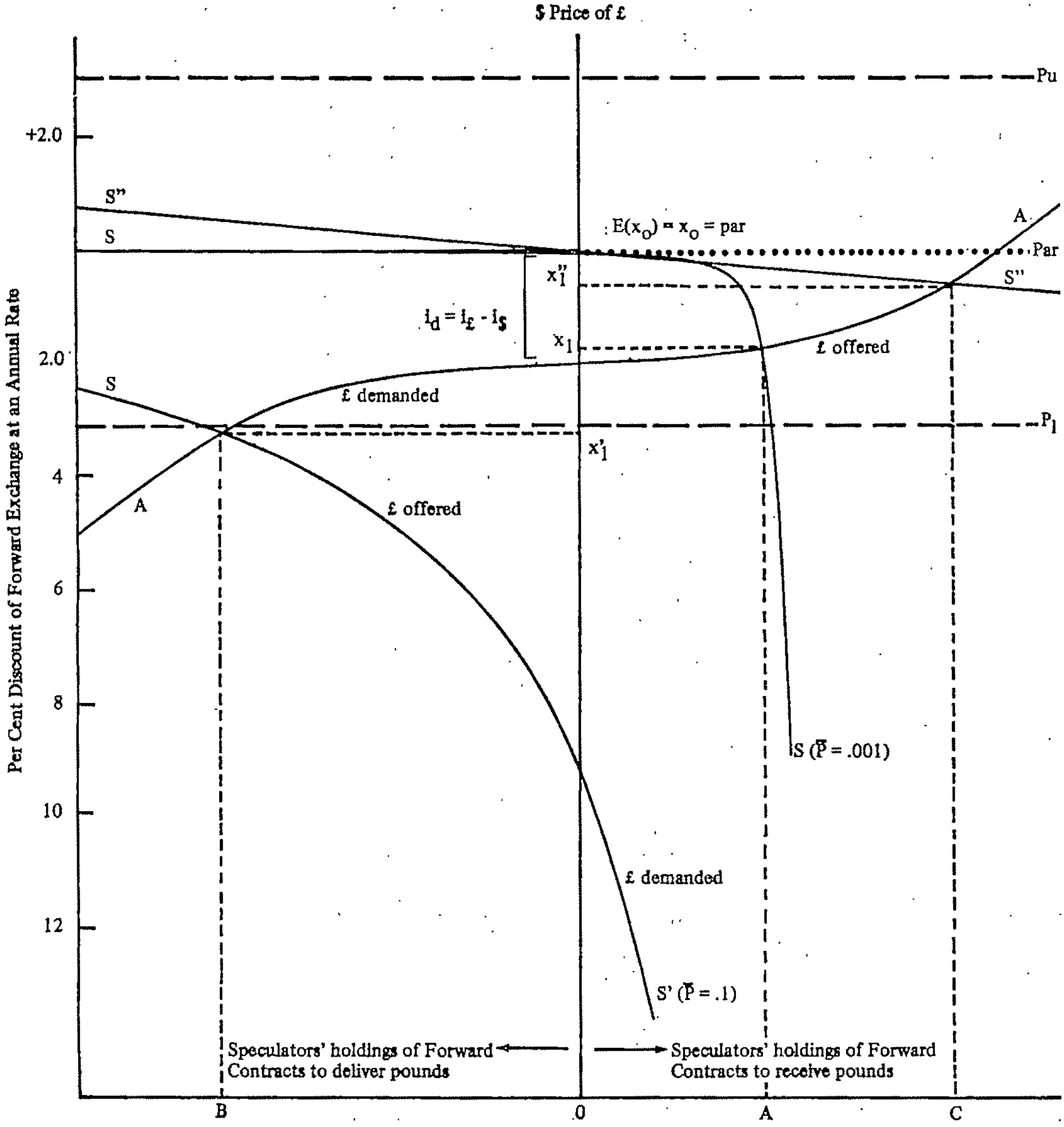


FIGURE 1.

interest earning assets held abroad under forward cover are not perfect substitutes for either interest earning assets or demand deposits in domestic currency. Hence, the elasticity of the *AA* curve may be expected to be a diminishing function of the covered yield on arbitrage funds.[12]

Equilibrium in the model is not appreciably affected by small changes in the values specified for x_2 and x_3. However, the model is highly sensitive to changes in the value of $\overline{P}$. Devaluations of the currencies of developed countries are infrequent. The American dollar has been devalued only once in the last one hundred years; the British pound has been devalued only three times. Even the French franc had remained stable for eleven years preceding the August 1969 devaluation. As a result, in the absence of any evidence

[12] If the model is broadened to include market participants other than pure speculators and pure arbitragers, the shape of the *AA* curve must be justified on more general grounds than those given here. See Grubel (pp. 12–21) and Schilling.

indicating that there will be unusual pressure on the par value of a currency during the immediate future, the level of speculative doubt represented by $\overline{P}$ is likely to be based upon the ability of a sovereign nation to devalue its currency rather than upon any perceived likelihood of its doing so. Speculators believe that the probability of devaluation is not quite zero, but is very small. This might be considered the usual or "normal" situation in present day exchange markets. It is represented in Figure 1 by intersection of the speculative excess demand curve SS ($\overline{P}=.001$) with the arbitragers' excess supply curve AA at forward rate x_1 with speculators' holdings of forward pounds (and arbitragers' holdings of spot pounds) equal to OA.[13]

The behavior of the model when in this state is essentially that predicted by the traditional *IRPT*. There is a positive speculative demand for the discounted forward currency which is quite price inelastic relative to the arbitrage supply of forward exchange. Therefore, the forward rate lies between the spot and parity forward exchange rates but quite close to the parity forward rate and will vary closely with the parity rate.

Occasionally something occurs which speculators can rationally interpret as indicating an unusually large possibility of devaluation in the near future. For example a new government may be coming to power whose attitude toward the maintenance of fixed exchange rates is unknown or suspect. Alternatively, an unusually large deficit in a country's balance of payments may show up suddenly or the government of the dubious currency may have recently followed inflationary domestic policies. Under these circumstances, $\overline{P}$ is quite likely to increase greatly in relative terms to a level sufficient to drop the speculators' excess demand curve for forward exchange well below the parity forward rate corresponding to any normal level of the interest rate differential.

For instance, if $\overline{P}$ were to increase from .001 to .1, the speculators' curve would shift from SS to $S'S'$. Its intersection with the arbitragers' curve would then occur on its elastic, forward-selling side and the forward pound would be driven to x_1'. Speculators will now resell their previous purchase of OA forward pounds and take a negative position, selling OB forward pounds to arbitragers who will shift assets to the low interest rate currency from the high interest rate currency until they have sold OB spot pounds. This situation is usually characterized by the term "crisis" or "speculative run."

The occurrence of a crisis as a result of devaluation fear is also a feature of the traditional *IRPT*. However, the equilibrium in the present model is sharply affected by increases in the value of $\overline{P}$ even when it remains small in absolute size. A subjective estimate of a one in ten probability of devaluation within ninety days is hardly a panic; yet it is sufficient to drop the SS curve to a position where its intercept with the vertical axis is more than 2 percent below the spot rate—well outside the intervention points. Furthermore, the SS curve is asymmetric, being more elastic on the left side than on the right side implying that speculators' pressure against a currency when it is under attack will be systematically greater than their previous support for it.

In order to see what happens when devaluation fear is completely absent let us assume that the risk of devaluation of any currency within three months is believed to be so small that it can safely be ignored. Speculators' subjective probability distribution of possible forward rates in ninety days will now be constrained by the spot rate support points.

[13] The intersection of the two curves determines the forward exchange rate at which speculation and arbitrage are balanced within this simplified model. The actual day to day position of the forward rate will reflect perturbations from this equilibrium resulting from the effects of 1) variations in the level and imbalances in the direction of flow demands, and 2) the presence of exogenous stock demands (inventory changes due to anticipations of goods price changes, strikes, etc.). Also the presence of transactions costs may cause lags in the adjustment of spot and forward positions, i.e., a speculator or abitrager who wishes to reduce his forward position may find that by waiting a couple of days and letting some of his presently held forward contracts mature he can reduce his costs by more than enough to offset the expected value of his loss from holding a non-optimal position meanwhile.

Let it be further assumed that the spot rate in ninety days is expected to be at either the upper or the lower support point with probability one-half; an assumption which gives the maximum risk consistent with a spot rate wholly constrained within its support points. The application of these assumptions to equation (7) yields

$$(9)\quad Y^* = .5x_1\left[\frac{1}{x_4 - x_1} + \frac{1}{x_5 - x_1}\right]K$$

Where

x_1, k and Y^* are defined as before and
x_4 = the lower support point
x_5 = the upper support point

Let it be assumed that the support points are the same as the present support points for the pound: 2.38/2.40 = 1.00833 times par for the upper support points, and 2.38/2.40 = .99167 times par for the lower support point. Equation (9) is then used to derive the speculative excess demand curve ($S''S''$) in Figure 1.

It is quite apparent that under a condition of negligible devaluation fear the forward position taken by speculators in the higher yielding currency would increase greatly (to OC in Figure 1) even though $S''S''$ was derived under assumptions which maximized the risk consequent upon movement of the spot rate within its support points. We conclude that in the absence of devaluation fear, the presence of even moderate international interest rate differentials would induce capital movements much larger than those which characterize the *IRPT* equilibrium today.

IV. *Concluding Remarks*

The recognition of the binomial character of speculative expectations which is embodied in our speculative excess demand function for forward exchange alters the traditional *IRPT* model in two ways: 1) the position of the speculative market schedule becomes highly sensitive to speculative expectations about devaluation; 2) the speculative market schedule is asymmetric in the presence of devaluation fear.

Insofar as the model presented is relevant to real world foreign exchange markets it suggests that: 1) The international monetary system in its present form contains its own gremlins, the frequency and ferocity of speculative attack upon a "weak" currency can be explained without invoking the existence of "gnomes"; 2) If the international monetary system were altered in a manner which assures that par values will remain permanently fixed (or will only slide slowly) without also widening the ban of spot rate variation, the quantity and sensitivity of international capital flows to interest rate differentials would increase greatly.

REFERENCES

J. Black, "A Proposal for the Reform of Exchange Rates," *Econ. J.*, June 1966, *76*, 501–06.

S. Black, "Theory and Policy Analysis of Short-term Movements in the Balance of Payments," *Yale Econ. Essays*, spring 1968, *8*, 5–78.

P. Einzig, *The Theory of Forward Exchange*, London 1937.

M. S. Feldstein, "Uncertainty and Forward Exchange Speculation," *Rev. Econ. Statist.*, May 1968, *50*, 182–92.

H. G. Grubel, *Forword Exchange, Speculation, and the International Flow of Capital*, Stanford 1966.

P. B. Kenen, "Trade, Speculation, and the Forward Exchange Rate," in R. E. Caves, et al., eds. *Trade, Growth, and the Balance of Payments*, Chicago 1965, 143–69.

H. Latane, "Criteria for Choice Among Risky Ventures," *J. Polit. Econ.*, Apr. 1959, *67*, 144–55.

D. J. Schilling, "Forward Exchange and Currency Position," *J. Finance*, Dec. 1969, *24*, 875–85.

H. R. Stoll, "An Empirical Study of the Forward Exchange Market Under Fixed and Flexible Exchange Rate Systems," *Can. J. Econ.*, Feb. 1968, *1*, 55–78.

S. C. Tsiang, "The Theory of the Forward Exchange Market," *Int. Monet. Fund Staff Pap.*, Apr. 1959, *7*, 75–106.

L. B. Yeager, *International Monetary Relations*, New York 1966.

On the Measurement of Import Substitution

By SAMUEL A. MORLEY AND GORDON W. SMITH*

Hollis Chenery, in his "Patterns of Industrial Growth" (pp. 639–44), introduced the measure of import substitution which has since become most widely used.[1] For Chenery, import substitution occurs with a decline in the ratio of imports to total supply. The magnitude of import substitution between two periods o and t is then given by:

$$(1)\quad IS_i = (\overset{o}{m}_i/\overset{o}{z}_i - \overset{t}{m}_i/\overset{t}{z}_i)\overset{t}{z}_i \qquad \begin{aligned} m_i &= \text{imports} \\ z_i &= \text{total supply of the products of sector } i \end{aligned}$$

Most authors have divided this number by the increase in domestic production to obtain the percentage of growth in the sector "accounted for," "due to," or "attributable to" import substitution.[2]

Because of its narrow definition of imports, however, Chenery's original approach misses much of what can be meaningfully considered as import substitution. In this paper we examine the nature of this bias and attempt to correct it by expanding our concept of imports to include the intermediate demands generated by import substitution. We then compare our measure to one proposed by Chenery, Shuntaro Shishido, and Tsunehiko Watanabe which included intermediates in a rather different fashion. Finally, we present calculations for Brazil 1949–64 to give some idea of the magnitude of the bias of the original Chenery measure for a semi-industrialized *LDC*. These results, together with a brief examination of import substitution in other countries, suggest that for long periods of industrialization and/or for international cross sections, import substitution measures which do not incorporate intermediate demands can be very misleading even as descriptive tools.

I. *The Concept of Imports and Total Supply*

The basic identity for the measurement of import substitution is:

$$(2)\quad x_i + m_i \equiv f_i + \sum_j a_{ij}x_j \qquad i = 1 \ldots n$$

m_i = imports
x_i = gross production
f_i = final demand, both domestic and foreign[3]
a_{ij} = observed input-output coefficient so that
$\sum_j a_{ij}x_j$ = total intermediate use of sector i's production

That is, total gross supply equals total gross demand in each sector.

Now, imports usually have been defined as m_i and total supply as $x_i + m_i$.[4] Thus, imports supplement domestic production in satisfying gross demand.

Alfred Maizels (p. 522) finds this definition inadequate—it does not value imports and domestic production on a "comparable basis." Whereas imports may be used *in toto* outside the sector, a portion of domestic production (a_{ii}) must be reserved to produce itself. Hence, in calculating total supply and demand Maizels eliminates intrasectoral

* Assistant professors at the University of Wisconsin and Rice University, respectively. This paper grew out of our participation in the Brazil Development Assistance Program of the University of California, Berkeley. We wish to thank Joel Bergsman, Donald Huddle, Ronald Soligo, and Richard D. Young for their comments and encouragement.

[1] Among those studies which have employed the Chenery measure are Lewis and Soligo, Steuer and Voivados, Ahmad, Huddle, Diaz-Alejandro, and Eysenbach.

[2] As a rule, import substitution of the Chenery type has been developed as one factor in a more general partition of the growth in output. Hence the division by the magnitude being "explained". The interested reader should see Chenery or Lewis and Soligo.

[3] For ease of presentation we aggregate the several components of final demand. This does not affect import substitution.

[4] See Chenery, Steuer and Voivados, Lewis and Soligo, Ahmad, Huddle, Diaz-Alejandro, Eysenbach.

sales from gross production (pp. 522–23).

Neither set of definitions recognizes that an import ultimately substitutes for or supplements the output of *many* domestic sectors. If an import is to be replaced without induced rises in imported inputs or reductions in the supplies available for final demand in other sectors, production must be increased not only in the industry finally processing the good, but also in its supplier industries and in their supplier industries and so forth.

In effect, the newly required intermediate output was previously "supplied" by the import of the final product. An accurate assessment of the total supply of each sector's products should therefore include such "implicit" imports. Only if this is done, will the two components of total supply, imports and domestic production, be measured on the same (gross production) basis.

Implicit imports are not included in the Chenery-Maizels (C-M) definitions, nor is their displacement by domestic production counted as import substitution. For example, suppose a country imports all its refined petroleum products in the initial year. Then oil is discovered and domestic refineries are built which process the domestic crude. Assume that final demand is constant. C-M will show import substitution in refining equal to the original imports, but none in crude oil. But the elimination of refined imports without importing crude was made possible only by this domestic production of crude. It, too, should be counted as import-substituting production.

These implicit or indirect imports can be incorporated with a sufficiently detailed and accurate domestic input-output table. Let A be such a table, and assume that the a_{ij} are constant over the relevant range. Then, in matrix notation,

$$(3) \qquad [I - A]x + m = f$$

$$(4) \qquad x + [I - A]^{-1}m = [I - A]^{-1}f$$

$$(5) \qquad m^* = [I - A]^{-1}m = \text{the vector of redefined imports}$$

$$(6) \qquad z^* = x + m^* = \text{the new vector of total supply}$$

$$(7) \qquad IS_i^* = (\overset{o*}{m_i}/\overset{o*}{z_i} - \overset{t*}{m_i}/\overset{t*}{z_i})\overset{t*}{z_i}$$

m^* converts imports to a gross production basis and allocates them to their proper domestic sectors. It can be viewed as the domestic production necessary to substitute completely for imports, holding all final demands constant. Complete substitution need not be feasible within any arbitrarily short period of time. Rather we assume only that if import substitution of any product does occur, the technology employed is accurately described by the a_{ij}.

II. *A Previous Attempt to Include Intermediates*

Chenery, Shishido, and Watanabe (CSW) incorporated intermediates into the analysis several years ago, but only by introducing a somewhat artificial definition of imports, different in its interpretation from both the traditional approach and the one set forth in the previous section.

To show this, we begin with the equations,[5]

$$(8) \qquad \overset{t}{x_i} = \overset{t}{f_i} - \overset{t}{m_i} + \sum \overset{t}{a_{ij}}\overset{t}{x_j} \qquad i = 1 \ldots n$$

$$(9) \qquad \overset{o}{x_i} = \overset{o}{f_i} - \overset{o}{m_i} + \sum \overset{o}{a_{ij}}\overset{o}{x_j}$$

f_i and m_i are considered to be exogenous, and for ease of exposition assume that $a_{ij}^t = a_{ij}^o$.[6] Multiplying (9) by $\lambda = GNP^t/GNP^o$, subtracting (9) from (8) and solving for x_i yields

$$(10) \qquad \overset{t}{x_i} - \lambda \overset{o}{x_i} = \sum r_{ij}(\overset{t}{f_j} - \lambda \overset{o}{f_j}) + \sum r_{ij}(\lambda \overset{o}{m_j} - \overset{t}{m_j})$$

where r_{ij} is the ijth element of the Leontief inverse.

With (10), CSW (pp. 106–07) partition non-proportional growth into exogenous "causes". The first term is the influence of non-unitary income elasticities of final demand, while the second defines the effects of import substitution.

[5] Again, for ease of exposition we aggregate final demands. CSW distinguish exports from domestic final demands.

[6] Differences in intermediate demands using the two matrices are identified as technological change by CSW (p. 106).

The use of $(\lambda m_j^o - m_j^t)$ is unusual, since it equates import substitution with the effects on sector i's production of changes in the (m_j/GNP).[7] Such changes arise both from modifications in import ratios as usually defined and from non-proportional expansion in the demand for industry j's products. As a result, a part of Chenery's import substitution in rapidly growing sectors appears as an income effect under CSW's criterion, while the contrary is true for slow growers. This, CSW themselves point out on page 118 in their article.

In the extreme, imports could completely absorb the market in slow growth sectors, while CSW showed positive direct import substitution.[8] In very high growth industries, domestic production could greatly increase its market share, with negative direct CSW import substitution.[9] These peculiarities partly explain why the CSW methodology has not been followed in subsequent studies. CSW include intermediates but do not preserve the original notion of import substitution as a decline in the ratio of a sector's imports to the total supply of its products.

Our approach does this and can be fitted into equation (10) as follows. Introduce another set of constants, P_{ij}. Subtracting and adding $\sum r_{ij}[(\lambda - P_{ij})m_j^o]$ to (10) and regrouping gives

$$x_i^t - \lambda x_i^o = \sum r_{ij}[(f_j^t - P_{ij}m_j^o) - \lambda(f_j^o - m_j^o)] + \sum r_{ij}(P_{ij}m_j^o - m_j^t) \tag{11}$$

The P_{ij} can be any reasonable projectors of imports.[10] Import substitution then becomes the difference between actual production and what production would have been had imports grown at the rates P_{ij}.

In our measure, all $P_{ij} = \sum r_{ij} f_j^t / \sum r_{ij} f_j^o = z_i^{*t}/z_i^{*o}$, for each sector i. Substituting in (11) and regrouping yields,

$$x_i^t - \lambda x_i^o = (z_i^{*t} - \lambda z_i^{*o})(1 - m_i^{*o}/z_i^{*o}) + z_i^{*t}(m_i^{*o}/z_i^{*o} - m_i^{*t}/z_i^{*t}) \tag{12}$$

The last term is IS_i^* (see equation (7)) while the first gives the effects of non-proportional growth in final demands. Equation (11) reduces to the type of partition Chenery used in "Patterns" (pp. 639–41) and followed by others (e.g., Ahmad, Diaz-Alejandro, Eysenbach, Huddle, and Lewis and Soligo).[11] Now, however, imports and total supply are defined to include the intermediates implicit in the import bill. The autonomous final demands and import ratios determine domestic product through the input-output matrix.

Although these open Leontief partition models help to clarify concepts, their limitations are clear. First, causation can be inferred only if the economy is as simple as the model. Furthermore, there is no a priori theoretical interpretation of declines in the import shares (however defined). Under certain restrictive assumptions (see Chenery, p. 628), changes in the *traditional* import ratios signify shifts in comparative advantage.[12] Unfortunately, actual variations in import ratios arise from many sources, rang-

[7] $\sum r_{ij}(\lambda m_j^o - m_j^t) = GNP^t[\sum r_{ij}(m_j^o/GNP^o - m_j^t/GNP^t)]$

[8] $(\Delta x/x < \Delta m/m < \Delta GNP/GNP)$

[9] $(\Delta GNP/GNP < \Delta m/m < \Delta x/x)$

[10] P_{ij} involving x_j are confusing. For example, let $P_{ij} = x_j^t/x_j^o$. Then the last term of (11) becomes $\sum r_{ij}[x_j^t(m_j^o/x_j^o - m_j^t/x_j^t)]$, the Chenery import substitution multiplied by the appropriate row of the inverse. This *does not* measure, however, the impact of changes in the m_j/z_j. Rather, it is the difference between x_i^t and what production would have been had the m_j grown at the same rate as the *observed* z_j, a totally different matter. It can be shown that if $x_j^t/x_j^o > m_j^t/m_j^o$, the change in import ratios implied by the use of this projector will be *greater* than those which actually occurred.

[11] If $\lambda = 1$, total growth in production is being partitioned. Import substitution is unaffected by the value of λ.

[12] Intermediate demands may then be included as follows. Define import substitution as the effect on production of changes in the ratio of imports to domestic production. Let S^t be a diagonal matrix with m_i^t/x_i^t on the diagonal. Then

$$x^t = Ax^t + f^t - S^t m^t$$

and

$$x^t - \lambda x^o = [I - A + S^o]^{-1}(f^t - f^o) + \{[I - A + S^t]^{-1} - [I - A + S^o]^{-1}\}f^t$$

The last term is the import substitution vector. Import substitution here acts *directly* on the inverse, and the effects of changes in any given m_i/x_i cannot be isolated.

ing from government policies to growth itself. Whatever the definition of imports, proportionality is but a convenient benchmark, much as the savings ratio, output per man-hour, etc.

In any real economy, then, import substitution is the *ex post* difference between actual imports in some period t and what imports would have been had import ratios remained at the levels of the base period.[13] Although our approach can be incorporated into a partition model, it is more aptly viewed as a method of measuring domestic production and imports on a comparable basis.

III. *Differences Between the Measures*

IS_i^* may be larger or smaller than IS_i (the traditional measure).[14] Note that

$$(13) \qquad (\overset{o}{m}_i/\overset{o}{z}_i-\overset{t}{m}_i/\overset{t}{z}_i)=(\overset{t}{x}_i/\overset{t}{z}_i-\overset{o}{x}_i/\overset{o}{z}_i)$$

Thus,

$$(14) \qquad IS_i^*-IS_i=(\overset{t}{x}_i/\overset{t*}{z}_i-\overset{o}{x}_i/\overset{o*}{z}_i)\overset{t*}{z}$$
$$-(\overset{t}{x}_i/\overset{t}{z}_i-\overset{o}{x}_i/\overset{o}{z}_i)\overset{t}{z}_i$$

$$(15) \qquad =(\overset{t}{z}_i/\overset{o}{z}_i-\overset{t*}{z}_i/\overset{o*}{z}_i)\overset{o}{x}_i$$

Equation (15) may be written as:

$$(16) \qquad (IS_i^*-IS_i)/\overset{o}{x}_i=\overset{t}{z}_i/\overset{o}{z}_i$$
$$-[\overset{t}{z}_i+(\overset{*t}{m}_i-\overset{t}{m}_i)]/[\overset{o}{z}_i+(\overset{*o}{m}_i-\overset{o}{m}_i)]$$

$IS_i^*-IS_i>0$ if, and only if, z_i, *direct* supply grows more rapidly than $(m_i^*-m_i)$, the *indirect* supply embodied in imports. Furthermore, the greater the difference in these growth rates, the greater the bias implied by the traditional approach.

The Chenery definitions are likely to yield underestimates for most semi-industrialized countries in recent decades. The process of import substitution in *LDC*'s seems to pass through a sequence of stages or dominant sectors:[15] 1) light consumer goods based on local raw materials; 2) consumer durables; 3) some intermediates and capital goods.

Of particular interest is the magnitude of import substitution as a country moves from light industry to more diversified manufacturing, i.e., from somewhere in stage 1) or early in stage 2) to somewhere in stage 3). Not only is this transition of great importance for the ultimate success of industrialization, but in addition all empirical studies of which we are aware have in fact measured import substitution between stages 1), 2), and 3). What differences would we expect in the results obtained from the two approaches?

For sectors producing exclusively for final demand, $IS_i^*=IS_i$, since indirect imports are always zero. Similarly, if the sector produces but little for intermediate use, the relative differences will be small. Clothing, transportation equipment, food processing, etc. fit in here.

At the opposite extreme, consider a pure intermediate industry, one for which $\sum_j a_{ij}x_j=x_i+m_i$. Total direct supply is a unique function of the production of the user sectors, while indirect supply is a linear function of imports. If import substitution in the traditional sense is occurring in the major users of such an intermediate product, then $IS_i^*>IS_i$, since $(\sum_j a_{ij}x_j^t/\sum_j a_{ij}x_j^o)>[(\sum_j r_{ij}m_j^t-m_i^t)/(\sum_j r_{ij}m_j^o-m_i^o)]$.

Some intermediates—metals is a prime example—are tied primarily to consumer durables and capital goods. In the stages we are considering, consumer durables imports grow much more slowly than their domestic production, often at a negative rate. If IS_i is not highly negative in capital goods and the intermediate itself—and none of the studies cited show this negative substitution for

[13] Eysenbach has pointed out that the *independent* influence of import substitution on output must be $z_i^o(m_i^o/z_i^o-m_i^t/z_i^t)$ rather than the usual $z_i^t(m_i/z_i^o-m_i^t/z_i^t)$. $z_i^t=z_i^o+\Delta z_i$, and Δz_i, the growth in total demand (=supply), is *itself* considered a "source of growth." Part of growth must therefore be explained by an interaction term, $\Delta z_i[\Delta m_i/z_i)]$. But as an *ex post* measure, the usual definition employing end-period total demands is perfectly valid. We must recognize, however, that import substitution itself has two "sources": changes in the import ratios and changes in total demands.

[14] Henceforth, variables with asterisks refer to the traditional Chenery definitions.

[15] See *inter alia*, Hirschman, Diaz-Alejandro and Maria da Conceição.

extended periods—the traditional measure will understate import substitution, perhaps, substantially so.

Other intermediates such as textiles, supply industries which absorbed most of the domestic market early in the industrialization process. Import substitution in this group of intermediates will not differ greatly for the two measures, since indirect imports will be a very small proportion of the total supply.

Finally, IS_i must always be zero for intermediates such as power and internal transportation, which are not traded internationally. However, $IS_i^*>0$, if their output growth rate exceeds that of their indirect imports. This usually occurs in the stages we are examining.

A particular country may not fit this pattern well, frustrating our attempt at broad generalizations. Only an examination of the basic underlying data can determine the direction of the bias. We return to this in Section V below.

V. *Brazilian Import Substitution, 1949–64*

Brazil in the period 1949–64 serves as a good illustration of the biases in the usual measure. In the period following World War II appreciable import substitution occurred in manufacturing. Not only was the import share in consumer durables and even capital goods drastically reduced, but by the early 1960's the major proportion of the resulting intermediate demands was supplied by domestic factories. Thus 1949–64 encompasses substantial portions of stages 2) and 3).

Table 1 presents calculations for this period using the two sets of concepts. Import substitution is expressed as a percentage of the growth in production, while sectoral figures were aggregated using value-added weights.[16] All series are in the prices and exchange rates of 1959, the year for which the 32×32 input-output table was computed. An explanation of data sources and procedures are in the Appendix.

Our definitions show about a third again as much import substitution as Chenery's in manufacturing and 53 percent more for the economy as a whole. Most of the differences in manufacturing, 87 percent, are accounted for by seven intermediate goods industries,[17] while the two approaches give rather similar results for industries producing mainly for final demand. Note also that our measure yields positive import substitution in such sectors as energy and transportation which are not import-competing industries.

Metals illustrate particularly well the differences in the two methods. In 1949, domestic production already satisfied almost 80 percent of the direct use of metals. But most of the direct supply of such heavy metal-using sectors such as transportation equipment, machinery and electric equipment was still imported (79, 64, and 82 percent, respectively). When import substitution occurred in these products, most of the resulting increment in the use of metals as intermediates was supplied by domestic industry. Our measure identifies this as import substitution in metals, the traditional measure, as an increase in domestic intermediate demand. The former is clearly more indicative of the actual process of growth.

The largest relative differences occur in mining, a primary intermediate, where $IS^*/\Delta X=.14$ and $IS/\Delta X=-.73$. This arises from the behavior of crude oil (included in mining) and refined petroleum products. In 1949, no crude oil was imported directly while practically all refined products were supplied from abroad. By 1964, however, more than 90 percent of the latter was produced domestically, while over 70 percent of

[16] Specifically, aggregate import substitution for sectors 1 through n is obtained as

$$\sum_{i=1}^{n}(IS_i/\Delta x_i)(v_i^t\Delta x_i)\Big/\sum_{i=1}^{n}v_i^t\Delta x_i)$$

where v_i = value-added/gross production in sector i. This is equivalent to

$$\sum_{i=1}^{n}v_i^t IS_i\Big/\sum_{i=1}^{n}v_i^t\Delta x_i,$$

which is the sum of import substitution in value-added terms divided by the total increase in value-added, assuming a fixed value-added ratio.

[17] Metals, lumber and wood, paper, rubber, chemicals, plastics, and textiles.

Table 1—Brazilian Import Substitution According to Different Measures[a] 1949–64

	1959 Cruzeiros					
	Our Measure			Standard Measure		
	$IS^*/\Delta X$	x/s^* 49	x/s^* 64	$IS/\Delta X$	x/s 49	x/s 64
Manufacturing	.440			.342		—
Non Met. Min.	.262	.788	.953	.208	.838	.972
Metals	.511	.463	.744	.123	.793	.872
Machinery	.722	.350	.749	.711	.359	.755
Elec. Eq.	.859	.156	.891	.835	.180	.907
Transp. Eq.	.841	.183	.929	.813	.211	.941
Wood	.122	.891	.956	.014	.986	.994
Furn.	.030	.975	.995	0	1.000	1.000
Paper	.332	.680	.872	.143	.864	.959
Rubber	.131	.887	.982	.036	.972	.998
Hides	.270	.906	.971	.217	.929	.981
Chemicals	.778	.219	.789	.686	.310	.856
Drugs	.260	.783	.935	.250	.793	.941
Perfume	.007	.993	.997	.002	.997	.999
Plastics	.496	.509	.986	0	1.000	1.000
Textiles	.123	.922	.986	.070	.961	.999
Clothing	.003	.998	.999	0	1.000	1.000
Food	.043	.943	.966	.034	.952	.971
Drinks	.191	.894	.984	.173	.911	.993
Tobacco	—	1.000	1.000	—	1.000	1.000
Print. & Pub.	.094	.891	.953	.024	.963	.979
Misc.	.312	.715	.881	.302	.724	.886
Agriculture	.024	.907	.917	.013	.949	.944
Energy	.181	.815	.922	0	1.000	1.000
Commerce	.036	.964	.983	0	1.000	1.000
Services	.056	.945	.965	0	1.000	1.000
Mining	.141	.330	.368	−.734	.712	.461
Constr.	.014	.985	.993	—	1.000	1.000
Transport.	.031	.964	.984	—	1.000	1.000
Whole Economy	.211			.138		

Source of data. See Appendix.
[a] Sectoral percentages are aggregated using value-added weights.

the crude oil was imported.[18] This resulted in large negative direct IS in mining, which was more than offset, however, by the positive indirect substitution from producing at home almost 30 percent of the crude oil which had 1949 conditions continued would have been imported in already processed form.

VI. *Evidence from Other Countries*

We have not made similar comparisons for other countries. However, we expect the bias to be significant in many other cases. Its direction and general magnitude for any intermediate product, can be determined by comparing two growth rates, those of direct supply and imports of the principal user sectors. If the first is substantially greater than the second, then the bias in the traditional measure will be considerable (see equation (16)). The data presented by Ahmad for India, Diaz-Alejandro for Argentina, CSW for Japan, and by Chenery in his "Patterns" all point to a significant underestimate in intermediates for the periods they cover. IS_i in all final and intermediate sectors is positive and appreciable. Thus, it is likely that IS_i^* was significantly greater than the standard measures in all intermedi-

[18] *Ministério de Planejamento* (pp. 45 and 55), W. Baer (p. 283).

ate sectors with indirectly imported supply.

Among published studies, the only possible exception to this pattern is Lewis and Soligo's on Pakistan. Even so, substitution in intermediates and capital goods as a whole and in most of their sub-categories is significant 1954–55 to 1963–64, the longest period covered by their estimates.[19]

VII. *Limitations and Observations*

Granted the desirability of such an approach as we have set forth, there are still several limitations to using observed input-output tables for actual measurement of import substitution.

1) The two-digit level of aggregation is too great. For example, petroleum refining appears with caustic soda in the Brazilian chemical industry.

2) Some imports cannot be produced at home due to lack of natural resources, technology, etc. All the same, these imports will be "decomposed" exactly as the average for production which does occur in the sector. Ideally, these products should be treated separately since they are not (economically) substitutable. We were unable to do this in the Brazilian calculations.

3) The a_{ij} are not usually fixed over time. Non-constancy may arise from (a) changing output composition within 2-digit *SIC* industries, (b) shifts in relative prices, and (c) from marginal a_{ij}'s different from the average.

The third problem is not serious *if* the table used accurately reflects end-period technology. Then import substitution will still measure differences in domestic output for that year "due to" a decline in the import share, holding technology constant. The 1959 Brazilian table no doubt differs from the one we would have observed in 1964. But similar results are obtained for 1949–59, when import substitution is 39 percent of manufacturing growth under our definitions and 30 percent with the traditional approach.

All of the other difficulties make errors in the a_{ij} likely in any concrete application. The basic issue, however, is whether a more accurate picture of import substitution can be obtained by letting all $a_{ij}=0$, as in the traditional approach, or by accepting erroneous, but observed $\hat{a}_{ij}$. It is barely conceivable that the $\hat{a}_{ij}$ leads to results farther from the correct a_{ij} than the traditional approach. In the input-output approach the bulk of the errors are likely to arise from aggregation problems. Without a more detailed table, they are difficult to evaluate.

VIII. *Conclusion*

In this paper we have introduced the implicit importation of intermediates directly into a Chenery-type measurement of import substitution, and we have shown how the traditional definitions, which ignore these linkages, miss import substitution in intermediate goods industries. If few intermediates are actually produced in a country, the differences in results will not be great. But if an accurate and fairly detailed input-output table is available, our measure will give a much more meaningful picture of import substitution for a country with well-developed intermediate industries.

Appendix

The input-output table was constructed by Willy van Rijckeghem using 1959 census data (pp. 388–94). We inverted it. Imports in 1959 dollars were grouped using Brazil's *IBGE* industrial (similar to the *SIC* two-digit) classification, as are the sectors of the input-output table. Relative prices in 1959, and hence the a_{ij}, should reflect 1959 exchange rates and tariffs for the different sectors. Hence each year's imports were converted into cruzeiros using 1959 effective exchange rates, including tariffs. The basic commodity imports and exchange rate data for 1959 and 1964 were kindly supplied by *IBGE*'s Laboratório Estatístico. For 1949 we relied primarily on the good quantitative estimates in Economic Commission for Latin America (see Maria da Conceição), with

[19] A downward bias is probably present in Maizels' slightly different estimates of import substitution, since imports in all but a few countries have been growing much more slowly than domestic production. His data are too aggregated for completely certain conclusions (pp. 150–61).

some corrections for differences in coverage.

The value-added series starts with 1959 value-added as computed from the census by van Rijckeghem (pp. 388–94). This series is projected back to 1949 and forward to 1964 through indices of real output supplied by the Brazilian Planning Ministry's Instituto de Pesquisa Economica-Social Aplicado (*IPEA*). The 1964 indices are unofficial estimates made by *IPEA* itself, while the remaining two years were computed by the Fundação Getúlio Vargas.

REFERENCES

J. Ahmad, "Import Substitution and Structural Change in Indian Manufacturing Industry," *J. Develop. Stud.*, spring 1965, *5*, 350–79.

W. Baer, *Industrialization and Economic Development in Brazil*, Homewood 1965.

H. B. Chenery, "Patterns of Industrial Growth," *Amer. Econ. Rev.*, Sept. 1960, *50*, 624–54.

———, S. Shishido, and T. Watanabe, "The Pattern of Japanese Growth," *Econometrica*, Jan. 1962, *30*, 98–139.

M. da Conceição, "The Growth and Decline of Import Substitution in Brazil," *Econ. Bull. Lat. Amer.*, Mar. 1964, *9*, 1–59.

C. Diaz-Alejandro, *Essays on the Economic History of the Argentine Republic*, Homewood, forthcoming.

M. L. Eysenbach, "A Note on Growth and Structural Change in Pakistan's Manufacturing Industry 1954–1964," *Pakistan Develop. Rev.*, Spring 1969, *9*, 58–65.

A. O. Hirschman, "The Political Economy of Import Substituting Industrialization in Latin America," *Quart. J. Econ.*, Feb. 1968, *82*, 1–32.

D. L. Huddle, "Postwar Brazilian Industrialization: Growth Patterns, Inflation, and Sources of Stagnation," in E. N. Baklanoff, ed., *The Shaping of Modern Brazil*, Baton Rouge 1969, pp. 86–108.

S. R. Lewis, Jr. and R. Soligo, "Growth and Structural Changes in Pakistan's Manufacturing Industry, 1954 to 1964," *Pakistan Develop. Rev.*, spring 1965, *5*, 94–139.

A. Maizels, *Industrial Growth and World Trade*, Cambridge 1963.

M. D. Steuer and C. Voivados, "Import Substitution and Chenery's Patterns of Industrial Growth: A Further Study," *Economia Internazionale*, Feb. 1965, *28*, 47–77.

W. Van Rijckeghem, "An Intersectoral Consistency Model for Economic Planning in Brazil," in H. S. Ellis, ed., *The Economy of Brazil*, Berkeley 1969, pp. 376–401.

Ministério de Planejamento e Coordenação Econômica, Escritório de Pesquisa Econômica Aplicada, *Petróleo: Diagnóstico Preliminar*, Rio de Janeiro 1966.

The Optimum Lifetime Distribution of Consumption Expenditures: Comment

By KAN HUA YOUNG*

In a recent article in this *Review*, Lester C. Thurow (p. 326) argued that "... individuals will continue to dissave as their income rises until current income equals desired current consumption. Thus if it were possible to estimate the income at which an individual becomes a zero saver for each year of his life, it would be possible to determine his optimum life distribution of consumption." To implement this idea, he suggested two alternative definitions of zero saving. Based on these definitions, two optimum lifetime distributions of consumption expenditures were estimated for individuals between the ages of 20 and 80 (see Figure 2 on p. 328). The idea appears to be quite ingenious. Its usefulness is appealing because the estimation procedure is relatively simple. The procedure involves merely plotting a graph of zero-saving income levels for various age groups of individuals.

Thurow was quick to warn us that his idea may have several limitations. Aside from limitations of technical nature, he pointed out that his procedure assumes that the lifetime utility function has annual utility levels as its sole arguments. In addition, the annual utilities are assumed to be independent and additive. Despite these limitations, the suggested procedure for the estimation of the optimum lifetime distribution of consumption expenditures may be very useful if it does not involve other limitations of a more serious nature. This note attempts to explore the possibility of developing an alternative estimation procedure and to examine the validity of the suggested procedure.

The fact that the suggested procedure uses only those observations with zero saving at each year of age and discards all other observations with positive or negative saving is somewhat disturbing. This implies that the lifetime distributions of consumption expenditures of those individuals with positive or negative savings are not optimum. One can argue, as Thurow did, that their distributions of lifetime consumption expenditures, particularly those with negative savings, are subject to institutional constraints. We must realize, however, that those people with zero saving are also subject to the same constraints. For example, if the interest rate were higher (or lower) than the actual level, their savings may be positive (or negative) instead of zero. In any event, it seems desirable to develop an estimation procedure which uses all available observations instead of only some of the observations.

In general, we can suppose that the average propensity to consume (which is equal to the marginal propensity to consume, if we assume a constant average propensity) is determined by both age and the relative income position of a particular individual. This proposition is consistent with both the permanent income and the relative income hypotheses.[1] For convenience of discussion, let us suppose that individual consumption functions may be reasonably approximated by

$$\text{(1)}\quad \begin{aligned} C(i, t) &= \alpha(i, t) + (\beta_0 + \beta_1 s + \beta_2 s^2 + \beta_3 t + \beta_4 t^2) I(i, t) \\ &= \alpha(i, t) + \beta(i, t) I(i, t) \end{aligned}$$

where $C(i, t)$ and $I(i, t)$ are consumption and income of individual i, while s and t are variables representing the relative income

* The author is research economist at Mathematica, Inc., Princeton, New Jersey.

[1] See James Duesenberry and Milton Friedman for further discussion on these hypotheses. See also Young for a more general model incorporating both of these hypotheses and some other hypotheses.

position and the age of the particular individual. The parameters $\alpha(i, t)$ and $\beta(i, t) = \beta_0+\beta_1 s+\beta_2 s^2+\beta_3 t+\beta_4 t^2$ are assumed to be different for each individual in different relative income positions at various ages.[2] In practice, the subscript i may actually represent a group of individuals. This expression can be estimated empirically. The resulting estimate of $\beta(i, t)$ will show how the propensity to consume varies as the relative income position and the age of individuals change. To suggest, as Thurow did, that the observed consumption $C(i, t)$ or the prediction generated from expression (1) is not optimum, one must argue that the observed consumption $C(i, t)$ and/or the observed income $I(i, t)$, and thus the propensity to consume $\beta(i, t)$, are not "optimum."

It is not only convenient but also reasonable to define optimum in terms of long-run equilibrium, so that an analogous relationship to expression (1) for optimum consumption and optimum income may be written as

$$(2)\quad \begin{aligned} C^*(i, t) &= \alpha^*(i, t) + (\beta_0^* + \beta_1^* s + \beta_2^* s^2 \\ &\quad + \beta_3^* t + \beta_4^* t^2)I^*(i, t) \\ &= \alpha^*(i, t) + \beta^*(i, t)I^*(i, t) \end{aligned}$$

where the functional form is formally identical with expression (1). Expression (2) is not completely specified without defining the optimum consumption and the optimum income explicitly. For our purpose, these two unobservable variables may be specified as

$$(3)\quad I^*(i, t) = \sum_{\tau=0}^{\infty} \lambda(t)[1 - \lambda(t)]^{\tau} I(i, t + \tau)$$

and

$$(4)\quad \begin{aligned} C^*(i, t) &= [1/\mu(i)]C(i, t) \\ &\quad + [1 - 1/\mu(i)]C(i, t - 1) \end{aligned}$$

where all notations have appeared previously, except the new parameters $\lambda(t)$ and $\mu(i)$, which may be called the expectation coefficient and the adjustment coefficient respectively.[3] In principle, expression (2), together with the specifications (3) and (4), can be estimated empirically.

For convenience let us examine two special cases of expression (2) where $\lambda(t)=1$ and $\mu(i)=1$. Under assumption $\lambda(t)=1$, which implies that the optimum income is identical with the observed income, expression (2) becomes

$$(5)\quad C^*(i, t) = \alpha^*(i, t) + \beta^*(i, t)I(i, t)$$

where identity (3) with $\lambda(t)=1$ has been used. In this special case, the particular approach suggested by Thurow assumed $\alpha^*(i, t)=0$ and $\beta^*(i, t)=1$ for those individuals with zero saving.

Alternatively, under assumption $\mu(i)=1$, which implies that the optimum consumption is identical with the observed consumption, expression (2) becomes

$$(6)\quad C^*(i, t)=\alpha^*(i, t)+[\delta(i, t)\beta^*(i, t)]I(i, t)$$

where $\delta(i, t)=I^*(i, t)/I(i, t)$ is the ratio of the optimum income to the observed income. In this special case, the particular procedure suggested by Thurow assumed not only $\alpha^*(i, t)=0$ and $\beta^*(i, t)=1$ but also $I^*(i, t)=I(i, t)$ and thus $\delta(i, t)=1$ for those individuals with zero saving. Finally, it is clear that the justification for the particular procedure suggested by Thurow, in both special cases, seems to be hard to find. This should also be true for the general case of expression (2). Consequently, his results may be questionable.

REFERENCES

J. Duesenberry, *Income, Saving, and the Theory of Consumer Behavior,* Cambridge 1949.

M. Friedman, *A Theory of the Consumption Function,* Princeton 1957.

L. C. Thurow, "The Optimum Lifetime Distribution of Consumption Expenditures," *Amer. Econ. Rev.*, June 1969, *59*, 282–97.

K. H. Young, "Demand for Sugar in the United States: A Synthesis of Time Series and Cross Section Analyses," unpublished doctoral dissertation, Columbia Univ. 1969.

[2] The particular functional form of $\beta(i, t)$ as shown is not essential. Furthermore, the exact definitions of variables s and t are unnecessary for subsequent discussion.

[3] These specifications correspond to the usual formulations of the permanent income and the partial adjustment hypotheses respectively. For further discussion, see Young.

The Optimum Lifetime Distribution of Consumption Expenditures: Comment

By Brian Motley and Samuel A. Morley*

In a recent contribution to this *Review*, Lester Thurow pointed out that as a result of institutional restrictions in the capital market the actual lifetime pattern of consumption expenditures may diverge from the optimal pattern. He went on to argue, on the basis of cross-section data, that this divergence is *in fact* substantial and that younger households would prefer to dissave to a *considerably* greater extent than they are at present able. If true, such a conclusion would have a number of policy implications. Unfortunately, however, his paper contains conceptual errors which vitiate his argument and force us to render a verdict of "not proven" on his stated conclusions.[1]

I. *A Critique of Thurow's Discussion*

Thurow's theoretical argument consists of three propositions: (i) there are institutional constraints on dissaving; (ii) these constraints are not absolute, however, and hence if we observe a household which is neither saving nor dissaving, we may infer that its current income is equal to its optimal[2] consumption level; and (iii) if preferences are homothetic, the path of lifetime consumption which a typical household would choose in the absence of borrowing constraints may be estimated from observations of the consumption levels of the zero-saving households in each age-group. We wish to take issue with each of these propositions; in particular we shall show that proposition (ii) is in general false, and that proposition (iii) is true only under the extreme assumption that the lifetime income of the zero-saving households is the same in every age group.

The first point to note is that it is not dissaving per se which is limited, but rather sustaining an expenditure pattern which requires the holding of negative non-human wealth (i.e. debts exceeding assets). A household which owns large holdings of marketable assets will have no difficulty maintaining a consumption level in excess of current income so long as this happy state of affairs continues. However, we do not stress this point[3] since it is not crucial to the theoretical discussion.

Thurow's second proposition *is* crucial and, we believe, is in general false. The fact that one observes a household with a balanced budget does not necessarily imply that it would behave in an identical fashion in the absence of institutional constraints on borrowing. At any point in time, the consumption plan chosen by the household depends on its aggregate wealth.[4] If the household has encountered a binding limitation on borrowing in the past, its present debt liabilities will be below the optimal level and hence its total resources will be greater than they would have been had no such limitation existed.[5] As a result it will choose a consump-

* The authors are assistant professors of economics at the University of Wisconsin. They acknowledge gratefully the financial assistance of the National Institute of Mental Health Training Program, The University of Wisconsin, Madison. They are indebted to Eugene Smolensky for helpful comments on an earlier draft of this paper. They gratefully acknowledge the support of the Behavioral Sciences Branch of the National Institute of Mental Health.

[1] Thus we do not contend that his conclusions are necessarily false (though we suspect they are) but that they cannot be substantiated by the type of test which he proposes.

[2] Throughout this paper the word "optimal" is used to designate the level (or path) of a variable which the household would choose if it were faced with perfect capital markets throughout its lifetime.

[3] Thurow mentions this point but states it is not important empirically. It is not clear on what grounds he makes this judgment.

[4] It is worth noting that with imperfect capital markets, the *time pattern* of future income receipts as well as their current discounted value is relevant to the making of consumption plans.

[5] In addition, as Thurow recognizes in his remarks under (3) on p. 326, the amount of past debts may influence the current borrowing constraint.

tion plan involving a higher level of expenditure in every future time period.[6] Thus the fact that a borrowing constraint is not binding *in the present* does not imply that the household's behavior would remain unchanged in the absence of such constraints, since *current* behavior is also affected by constraints which operated *in the past.*

Thurow is not, of course, primarily interested in the spending behavior of zero-saving households. He is concerned with the optimal consumption pattern of the "average" household.[7] In order to discover this pattern, he wishes first to identify a set of households in each age group whom he may assume to be consuming optimally; he chooses the zero-savers for this purpose. He then "scales up" the consumption expenditures of the zero-savers to the level of the average household by appealing to the assumption that preferences are homothetic (that is, the lifetime distribution of consumption expenditures is independent of the level of lifetime income). However, his procedure is valid only under the unstated (and almost certainly false) assumption that the lifetime income of the *zero-savers* is the same in every age group.[8] In the absence of this assumption, his results are meaningless since they relate to the behavior of a non-homogenous group of households.

In order to concentrate on this point, assume that the consumption expenditures of the zero-savers in each age group are indeed optimal. Define:

C_t^* = consumption expenditures of the t-year old group of zero-savers.

W_t^* = lifetime income of the households who are zero-savers at age t.

C_t^0 = optimal consumption expenditures of the average household at age t.

$\overline{W}_t$ = lifetime income of the average t-year old household.

Assume the preferences of all households are identical and homothetic. Assume a stationary economy so that $\overline{W}_t = \overline{W}$ (a constant) for all t. Then since C_t^* is on the optimal consumption path of the households who have lifetime income W_t^* and since the assumptions of identical and homothetic preferences assure us that the optimal share of lifetime income which the average household would devote to consumption expenditures at age t is equal to the share which t-year old zero-savers will devote, we may write

$$C_t^0 = \frac{\overline{W}}{W_t^*} \cdot C_t^* \tag{1}$$

The practical problem, of course, is that the lifetime income of those households which are zero-savers at age t (i.e., W_t^*) is not observable.

Thurow adjusts the zero-saver's consumption expenditures differently. He estimates optimal consumption of the average household at age t as:

$$\hat{C}_t^0 = \frac{\overline{W}}{\sum_{s=1}^{T} C_s^*} C_t^* \tag{2}$$

where T is the number of age groups into which the data are classified. This procedure requires the assumption that the lifetime wealth of the zero-saving households is the same[9] in every age-group (i.e. W_t^* is constant for all t and hence equals ΣC_s^*). Thurow marshalls no theoretical or empirical arguments in support of this assumption. However, if this assumption is false his results are strictly meaningless since the so-called optimal consumption level relates to a different lifetime wealth in each age-group. In fact, the cross-section data suggest that even in a stationary world the lifetime wealth of,

[6] The argument requires the assumption that neither future nor present goods are inferior.

[7] Many definitions are possible of an average household. In this paper we define it as a household whose income at each age is equal to the mean of all households of the same age.

[8] Note that this is a much stronger assumption than that "individuals in older age brackets are similar in all relevant aspects with those in younger age brackets." See also fn. 9.

[9] In the context of a growing economy, the required assumption is that $\overline{W}_t/W_t^*$ is constant for all t. Incidentally, if this ratio *were* constant and if the pattern of income receipts of zero-savers were similar to that of the average household, no bias of the type suggested by Thurow in his assumption (2) p. 326, would arise.

for example, twenty-five year old zero-savers would almost certainly be higher than that of fifty-five year old zero-savers. In 1961 twenty-five year old zero-savers (by Thurow's first definition) had incomes close to the age group mean of \$5000. Hence it is likely that their incomes were close to their normal level (see Milton Friedman p. 34). If their income profile were similar to that of the average household, they could expect to earn close to \$6700 at age fifty-five if no growth occurred in the economy. The actual income of fifty-five year old *zero-savers* was much lower than \$4000. Although this probably contained a negative transitory component (since it was below the age group mean) it is inconceivable that their lifetime wealth was not also lower than that of the twenty-five year olds. Hence Thurow's conclusion that the former group wishes to consume more than they are able at present and the latter group less may reflect nothing more than the fact that he has chosen noncomparable households. If the lifetime wealth of the twenty-five year old zero-savers is greater than that of the fifty-five year old zero-savers ($W^*_{25} > W^*_{55}$), his estimate of optimal consumption will be biased upward for twenty-five year olds and downward for fifty-five year olds.

In using cross-section data to make inferences about optimal behavior over time, two criteria are required: (i) at any age the households picked as representative are likely to be consuming optimally, and (ii) the households chosen from different age-groups should be "equivalent," by which we mean that they should have the same lifetime wealth. Our discussion in this section suggests that zero-savers may satisfy neither of these criteria. Since past constraints influence current behavior, zero-savers are unlikely to be consuming optimally. Moreover, there seems no reason to believe that zero-savers in the population have the same lifetime income regardless of their age, even in a stationary economy.[10]

[10] Again we wish to point out that the assumption that average individuals in older age brackets "... are similar in all relevant aspects with those in younger age brackets ..." does *not* imply that the same is true of zero-saving households.

II. *An Alternative Model*

We now construct a simple model in which perfect capital markets (with no restraints on household behavior apart from the lifetime budget constraint) assure us that consumption patterns are optimal. We shall show that if the data generated by this model were analyzed by Thurow's method, the shape of the zero-savings path would differ from that of the consumption path of the average household. We require three additional pieces of notation:

Y^N_{ti} = normal or expected income of household i at age t.

Y^M_{ti} = measured income of household i at age t.

W_i = lifetime wealth of household i.

Normal income corresponds to the concept which Friedman (pp. 21 and 93) describes as the "permanent component" of measured income. Symbols with bars will denote averages over households of the same age. We make the following (strong) assumptions:

(i) Average measured income is equal to average normal income at any age (i.e. $\bar{Y}^N_t = \bar{Y}^M_t$).

(ii) Normal income is positively correlated with measured income. Specifically we assume normal income is related linearly[11] with measured income. Combining this with assumption (i) we obtain

$$Y^N_{ti} = (1 - P)\bar{Y}^M_t + PY^M_{ti} \tag{3}$$

where P is less than unity and for simplicity is assumed constant over all age groups.

(iii) All households have identical homothetic preferences and behave according to the permanent income hypothesis:

$$C_{ti} = k_t W_i \tag{4}$$

We make no distinction between measured and normal consumption.

(iv) All households have the same normal income "profile" though the "heights" of the

[11] This relation will be linear if normal income and transitory income are uncorrelated and are distributed according to a joint normal distribution. Under this assumption the coefficient P is the fraction of the total variance of income in any age group contributed by the permanent component.

profiles differ; that is, for any household, the ratio of its normal income to that of the average household is constant throughout its lifetime. This means that at any age the ratio of normal income to lifetime wealth is the same for all households:

$$(5)\qquad \alpha_t = \frac{Y^N_{ti}}{W_i} = \frac{\overline{Y}^N_t}{\overline{W}}$$

Combining (4) and (5) yields

$$(6)\qquad C_{ti} = \frac{k_t}{\alpha_t}\, Y^N_{ti} = \gamma_t Y^N_{ti}$$

In Figure 1 we show the cross-section relationships expressed in equations (3) and (6). At measured income level Y^*, consumption and measured income are equal. Substituting (3) into (6) yields:

$$C_{ti} = \gamma_t(1-P)\overline{Y}^M_t + \gamma_t P Y^M_{ti}$$

Define Y^*_t as the income level at which savings are zero. Then

$$(7)\quad Y^*_t = \gamma_t(1-P)\overline{Y}^M_t + \gamma_t P Y^*_t = \left[\frac{\gamma_t(1-b)}{\gamma_t - b}\right]\overline{Y}^M_t = \left[\frac{1-b}{\gamma_t - b}\right]\overline{C}_t$$

where $b=1/P$. We may use this relation to examine the path of Y^* (which Thurow wrongly treats as the optimal path) under various assumptions with respect to normal income and optimal consumption. A striking example is the case in which optimal consumption is constant over the lifetime. Setting $\overline{C}_t=\overline{C}$ and differentiating (7) with respect to age, we find that Y^* moves inversely to Y^N:

$$(8)\quad \frac{dY^*}{dt} = \frac{-(1-b)\overline{C}}{(\gamma-b)^2}\,\frac{d\gamma}{dt} = \frac{\gamma^2(1-b)}{(\gamma-b)^2}\,\frac{dY^N}{dt};\quad (1-b)<0$$

Thus if the normal income path is concave and the true optimal path is constant, Thurow's technique yields a *convex* path. This result is illustrated in Figure 2a. A second example is shown in Figure 2b in which optimal consumption grows to a peak earlier in the lifetime than does normal income. Thurow's method yields a path which rises and falls but later rises again. Finally Figure 2c depicts the path of Y^* if the observed paths of average income and consumption in the cross-section study used by Thurow are assumed to represent normal income and optimal consumption, and the value of b is taken as 1.25. The similarity between the shape of this path and that of Thurow's "optimal #1" is quite striking (see p. 328), thus providing evidence that a model which assumes perfect capital markets will explain the zero-saving path quite adequately.

Equation (7) suggests an alternative technique by which the optimal consumption path may be estimated from cross-section data. One may take observations on Y^*_t and $\overline{Y}^M_t$ and solve equation (7) for γ_t. Evaluating (6) at the mean income level will provide an estimate of the optimal consumption path of the average household. The limitations of this method are obvious in view of the stringent assumptions underlying (7). Moreover the results are sensitive[12] to the value of the

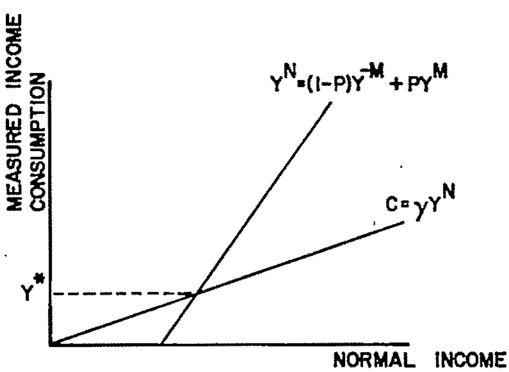


FIGURE 1. MEASURED AND NORMAL INCOME AND CONSUMPTION AT A GIVEN AGE.

[12] This can be seen by implicitly differentiating (7) for fixed Y^* and $\overline{Y}^M$:

$$\frac{d\gamma}{db} = \frac{\gamma(1-\gamma)}{b(1-b)} \begin{array}{ll} <0 & \text{for } \gamma<1 \\ >0 & \text{for } \gamma>1 \end{array}$$

If we over-estimate b, we will under-estimate optimal consumption of middle-aged households for whom γ is likely to be less than one.

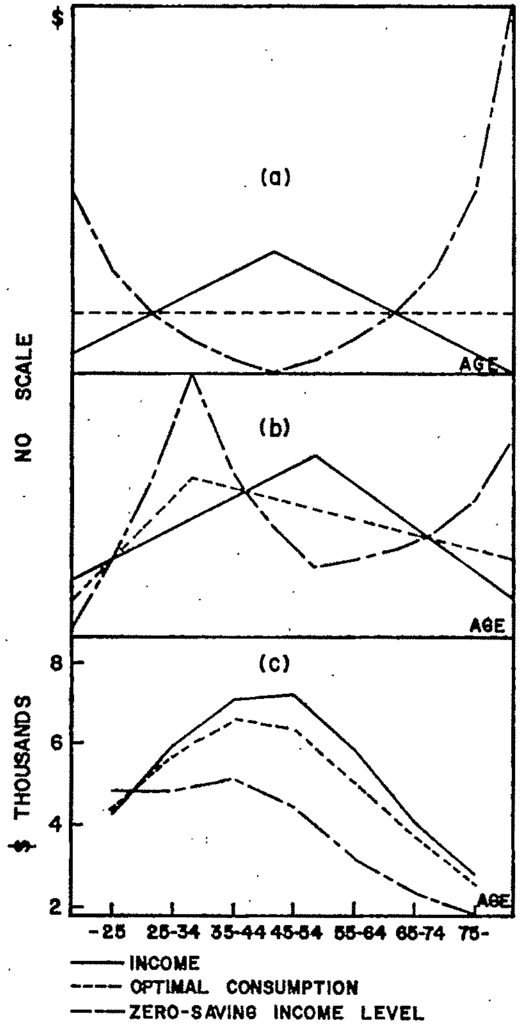


FIGURE 2. ZERO-SAVING INCOME LEVELS UNDER VARIOUS ASSUMPTIONS.

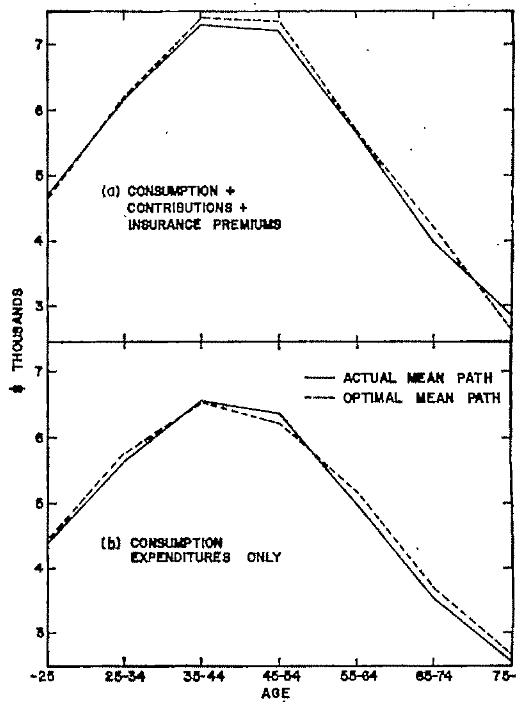


FIGURE 3. ACTUAL AND OPTIMAL CONSUMPTION PATTERNS 1961.

unknown parameter b. As is well known, however, (see Friedman p. 32), the parameter $P(=1/b)$ is equal to the ratio of the cross-section marginal propensity to consume to the average propensity and hence probably lies between .75 and .85. Figure 3 shows the path of actual consumption compared with the estimated optimal path under the assumption that $P=.80$ (i.e. $b=1.25$). They are obviously extremely close, suggesting that any divergence between actual and optimal consumption is quite small.[13] However, we wish to stress that this evidence is only suggestive and we fully recognize the limitations of our technique.

III. *Conclusions*

This paper has sought to show that there are serious errors in Thurow's paper which nullify his empirical results. His assertion that zero-savers are likely to be behaving optimally has been shown to be false. His "scaling up" technique has been shown to depend on a critical and probably false assumption. Finally we have developed a simple model which *can* be used to predict optimal consumption and in which the patterns of the zero-saving income level and of optimal consumption are necessarily different. Preliminary results suggest that the divergence between actual and optimal consumption, at least in the aggregate, is not particularly large.

[13] The reader should note that the close relationship between estimated optimal and actual mean consumption paths shown in Figure 3 and that between the estimated and actual zero-saving paths discussed earlier are not two independent demonstrations of the validity of our model.

REFERENCES

M. Friedman, *A Theory of the Consumption Function*. Princeton 1957.

L. Thurow, "The Optimum Lifetime Distribution of Consumption Expenditures" *Amer. Econ. Rev.*, June 1969, *59*, 324–31.

Bureau of Labor Statistics, *Survey of Consumer Expenditures, 1960–61. Consumer Expenditure and Income Cross-Classification of Family Characteristics, Urban United States,* Supp. 2, Part A to BLS Report 237–38, July 1964.

The Optimum Lifetime Distribution of Consumption Expenditures: Reply

By Lester C. Thurow*

I agree with Mr. Kan Hua Young that my calculations of the optimum lifetime distribution of consumption expenditures depend upon the existing interest rates (the interest rate appears in the first and only equation in the paper), that the individual must know, or at least have expectations about, his future income (the paper states that the individual has an expected lifetime budget constraint), and that the marginal utility of consumption expenditures may depend upon relative as well as absolute incomes (the paper does not address itself to what causes the marginal utility of consumption expenditures, but the empirical estimates are certainly compatible with his hypothesis).

The rest of his comment flows from his misconception that the article "implies that the lifetime distributions of consumption expenditures of those individuals with positive or negative savings is not optimum." Such an implication is not made, is not necessary, and the reverse is in fact assumed. The estimation procedure assumes that every individual with zero, positive, or negative savings is an optimum consumer given all of the constraints (current income, borrowing possibilities, lifetime income, etc.,) which he faces. The article attempts to determine what his optimum consumption would be if he faced only *one* constraint—his lifetime budget constraint.

The paper attempted to argue that if a person were neither dissaving nor saving, he had reached his desired consumption level given his personal discount differential, his consumption preferences, and his lifetime budget constraint. The assumptions behind this argument are outlined in the original paper and will not be repeated here. Let me repeat, however, that the paper does not assume that individuals with positive or negative savings are in any sense non-optimizing consumers.

The first section of the comment by Brian Motley and Samuel Morley lists some of the assumptions that were necessary to calculate empirically the optimum lifetime distribution of consumption expenditures. For some unknown reason they wish to charge that these are unstated assumptions which when revealed vitiate the analysis. In reality, each and every one of their list of assumptions was explicitly stated in the original article along with some other necessary assumptions (see pages 326 and 327). I am tempted to prove the point by quoting myself, but will let the reader reread the original assumptions if he is interested. Instead I will focus on the problem of whether the necessary assumptions are so strong as to make the empirical results useless.

There is no doubt that wealth allows one to dissave. Motley and Morely admit that theoretically this is not an important observation. Empirically it is not important as well, because, as assumption (3) mentions, individuals are increasing assets rather than reducing assets as the point of zero savings is approached. Empirically individuals are not dissaving out of wealth, they are borrowing resources to expand their consumption opportunities.

In assumption (3), I recognized that past constraints can influence current behavior and pointed out the direction of bias that this produced. Individuals would consume too little while young and too much while old in comparison with their optimal distribution. Empirically, the more severe the constraints on borrowing and the farther the actual income distribution is from the desired income distribution, the more important this bias becomes. The existence of the

* Massachusetts Institute of Technology.

bias, however, strengthens the central empirical conclusion that there should be more redistribution toward the young. If anything, my analysis underestimates the amount of consumption that individuals wish to do while young.

The assumption of homothetic preferences is important, but as footnote 14 indicated, the empirical results do not seem to depend upon this assumption. Two definitions of savings were used—one where the lifetime income of zero savers was less than the average lifetime income and one where the lifetime income of zero savers was greater than the average lifetime income. As stated, the results are approximately the same, regardless of whether consumption is scaled up or down.

In assumption (2) I recognized that the procedure requires that the lifetime income of zero savers be equal and pointed out the direction of bias that this assumption produces ("ignoring this factor biases the results toward too much consumption in the early years of an individual's lifetime"). This bias is partly offset by the other bias mentioned above, but there is certainly no reason to assume that they would be equal in magnitude. I have no idea which would be bigger.

If the estimates are to be explained by this factor, however, there must be a very peculiar distribution of lifetime income over different age groups. The lifetime income of a man 60 years old would have to be just half of the lifetime income of a man 40 years old. Assuming that economic growth leads to a real rate of growth of 2 percent per year in family income, the lifetime income of a 40 year old is 44 percent higher than that of a 60 year old. At the other end of the scale, the lifetime income of a man 20 years old would have to be less than the lifetime income of a man 40 years old to explain the estimated results. I can swallow the former, but not the latter.

Interestingly, the assumptions which were "not necessarily true" in the first section become have "been shown to be false" in the conclusions. The last time I studied logic, the former did not entail the latter.

The Motley-Morley conclusions that there is little divergence between actual and optimal consumption is not surprising, since this is precisely what they have assumed. Their estimating equation (Equation 7) is:

$$\overline{Y}_t^M = \frac{1-b}{\gamma_t - b}\,\overline{C}_t$$

where:

$(\gamma_t - b)/(1-b) =$ the average observed propensity to consume for each age group

$\gamma_t =$ the desired consumption propensity.

As long as "'b'' is judiciously selected, which it was, and as long as $(\gamma_t - b)/(1-b)$ is close to one, which it is, then γ_t will necessarily be almost equal to $(\gamma_t - b)/(1-b)$. Thus, they will necessarily find that actual consumption is very close to optimal consumption.

Tax Policy and Investment Behavior: Further Comment

By ROBERT EISNER*

There is now a rapidly accumulating body of work indicating that joint estimates by Robert Hall and Dale Jorgenson (1967, 1969), as to the magnitude and speed of effects of tax policy on investment are extremely high. Lawrence Klein and Paul Taubman (see Gary Fromm) have been more moderate on the basis of relatively unconstrained theoretical relations embedded in a multi-equation model. Charles Bischoff (see Fromm) is lower on magnitude and particularly speed of adjustment, with his stress of a putty-clay model which suggests much slower reaction to the relative price changes through which tax incentives for investment are usually seen to operate. Robert Coen (1969 and in Fromm) and Eisner (June 1969), along with Eisner and M. Ishaq Nadiri (1968) have pointed to substantially lower estimates of tax policy effects resulting from relaxation of the Hall-Jorgenson constraint that the relevant price elasticity is unity. The latter works have involved data either closely similar to or identical with those employed by Hall and Jorgenson and by Jorgenson and James Stephenson in other, related papers.

In their reply to the separate comments by Coen and by Eisner on their original "Tax Policy and Investment Behavior" (1967), Hall and Jorgenson: 1) cite findings from other studies to support a crucial but previously arbitrary assumption that the elasticity of substitution is unity; 2) dismiss the evidence by Coen, Eisner, and Nadiri that the price elasticity of capital demand found in unconstrained estimates of their own investment function is in fact very considerably less than unity, thus proportionately reducing the estimated effect of tax policy on investment; 3) develop new estimates of the lag pattern of investment which indicate a faster response to tax policy than does their earlier work; 4) estimate the impact of tax policy changes since 1963. The further estimates all depend upon the applicability of assumptions, particularly those implying unitary price elasticity, which have been called into question. I shall hence not endeavor to use this space to do more than observe that the new estimates by Hall and Jorgenson, as much as, or more than the old, are out of line, as to magnitude or speed of adjustment or both, with those of other investigators. To appreciate some of the magnitudes involved, the reader may be reminded of the tabulation prepared by Coen (1969, p. 374), showing some 6.7 billions of 1954 dollars of investment attributable to tax incentives from 1954 to 1963 under the Hall-Jorgenson constraint of unitary price elasticity, compared with a figure under 2 billion for the Coen-estimated overall elasticity of .2. Some of the Eisner-Nadiri and Eisner estimates would imply even lesser magnitudes of investment attributable to tax incentives. But the essential point is the extreme sensitivity of these magnitudes to the value of the price elasticity which is assumed or estimated.

I. *Other Estimates of the Elasticity of Substitution*

The defense now offered by Hall and Jorgenson for their assumption of unitary price elasticity may be welcomed as an acknowledgement of its crucial nature. This defense rests on certain extraneous and presumed favorable estimates of the elasticity of sub-

* Professor of economics, Northwestern University, and research associate, National Bureau of Economic Research. The author is indebted for extensive computational and related assistance to Betty Benson, Richard Strauss, and Robert Welch and to Tugrul Aladag, Joel Auerbach, Veena Bhatia, Dosier Hammond, Arnold Katz, Allan Mendelowitz, and John Soladay, to Charles W. Bischoff and Robert M. Coen for comments on an earlier draft, and to the National Science Foundation for financial support.

stitution which are tied to the further precarious assumption and role of perfect competition in the Jorgenson model.

I shall not attempt here a comprehensive review of all the published estimates of the elasticity of substitution, from a great variety of data and models. It would be judicious, at the least, to remind readers again that the prime characteristics of all of these various estimates, as is made clear in the impressive survey article by Marc Nerlove, is their considerable variance. There is hardly any consensus as to *the* elasticity of substitution, or indeed whether any reasonably constant elasticity of substitution exists. Estimates have been made, inevitably, on the basis of a number of critical simplifying assumptions such as perfect competition, homogeneity of factors and identical production functions across observations, to name just a few. One might reasonably view such estimates of the form and parameters of "the" production function as experimental rather than as firm building blocks for definition of an investment function.

What is more, although Hall and Jorgenson refer to them as "erratic" and "unreliable" (1969, pp. 391–92), estimates of the elasticity of substitution from time-series have been remarkably lower than those from cross-sections. Yet it is the time-series which might be expected to prove most directly relevant to estimates of the dynamic path of investment. Cross-sectional estimates, even if unbiased,[1] are more likely to relate to long-run, inter-equilibria elasticities. They may tell us little about the quarter by quarter and year by year adjustment to changes in relative prices which must be the stuff of the investment process.[2]

[1] Most recently Lucas has indicated a likely bias toward unity in cross-section estimates of elasticity of substitution and has suggested time-series estimates in the neighborhood of .4 as more reliable. (See Robert Lucas, p. 25; also Arnold Harberger, in Harberger and Bailey, pp. 7–8.)

[2] There is a substantial literature warning of pitfalls in the application to time-series relations of extraneous estimates of parameters from cross-sections. The reader may note, for example, the early work of Edwin Kuh, and recently, that of V. K. Chetty.

II. *Our Estimates of Price Elasticity*

In rejecting the direct estimates by Eisner and by Eisner and Nadiri of the effects of changes in relative prices, tax factors and output on investment expenditures, Hall and Jorgenson assert that the Eisner-Nadiri log linear model and the Jorgenson-Stephenson model (essentially the Hall-Jorgenson model) "are mutually contradictory and that evidence predicated on the validity of one is not relevant to the validity of the other," referring for support to Jorgenson and Stephenson (1969). Since the assertion is unexplained in the Hall-Jorgenson "Reply," the reader may best consult the Jorgenson-Stephenson reference and the Eisner-Nadiri reply to it for further evaluation. It is shown by Eisner-Nadiri (1970, Table 1) that the difference between arithmetic and logarithmic formulations, which Hall and Jorgenson claim makes the latter invalid, amounts to differences, for critical elasticities, between .197 and .201, .155 and .147, .058 and .052, .1314 and .1306, and .049 and .054.

Hall and Jorgenson (1969) charge that "Bischoff has drawn attention to an important error of specification in the econometric model used by Eisner-Nadiri" which "carries over directly to Eisner's model." They argue that:

1. "Errors in the distributed lag function employed in the model are autocorrelated . . . " and ". . . Eisner and Nadiri have ignored this autocorrelation so that their estimates of parameters of the distributed lag function are inconsistent."
2. "The estimated elasticity of substitution is highly sensitive to this error of specification."
3. Bischoff has demonstrated that elimination of the error in specification offers essential support to the Jorgenson model.

First, any "important error of specification" originates not with Eisner and Nadiri but with Jorgenson. We took over the Jorgenson specification along with the Jorgenson data expressly in order to focus as specifically

TABLE 1—MAXIMUM LIKELIHOOD ESTIMATION OF AUTOREGRESSIVE PARAMETER, JORGENSON-STEPHENSON DATA, MANUFACTURING STRUCTURES AND EQUIPMENT[a]

$$[1-\rho L]\left[\ln K_t = \gamma_0 + \sum_{i=1}^{7} (\gamma_{pi} \ln (p/c)_{t-i} + \gamma_{qi} \ln Q_{t-i}) + \sum_{j=1}^{2} \omega_j \ln K_{t-j} + v_t\right]$$

$$\hat{E}_p = \sum \hat{\gamma}_p/(1-\sum \hat{\omega}_j) \qquad \hat{E}_q = \sum \hat{\gamma}_q/(1-\sum \hat{\omega}_j)$$

$$t_{(1-\hat{E}_p)} = (1-\sum \gamma_p - \sum \omega_j)/\sigma_{\Sigma\gamma_p+\Sigma\omega_j} \qquad t_{(1-\hat{E}_q)} = (1-\sum \gamma_q - \sum \omega_j)/\sigma_{\Sigma\gamma_q+\Sigma\omega_j}$$

ρ (1)	$\sum e^2 \times 10^4$ (2)	$\hat{E}_p$ (3)	$\hat{E}_q$ (4)	$t_{(1-\hat{E}_p)}$ (5)	$t_{(1-\hat{E}_q)}$ (6)
A. Using c_1 as Rental Price of Capital					
0.0	.5698	.1729	.4493	1.19	2.00
0.3	.5437	.0929	.5554	1.72	1.94
0.6	.5527	.0470	.6334	2.55	1.80
0.9	.5217	.2131	.3404	4.15	2.78
0.925	.5149	.2017	.2908	4.21	2.88
1.0	.5796	.1470	.6889	2.56	.91
B. Using c_2 as Rental Price of Capital					
0.0	.5984	.1032	.7815	2.77	2.48
0.3	.5707	.0889	.7667	2.91	2.67
0.41	.5689	.0832	.7599	2.92	2.71
0.6	.5736	.0714	.7449	2.92	2.71
1.0	.6381	.0525	.5974	2.53	1.17

[a] As in previously published work, L denotes a lag operator, K = capital stock, p = price index of output, c = the rental price or "user cost" of capital, Q = output, $\hat{E}_p$ = estimated elasticity with respect to relative prices, p/c, and $\hat{E}_q$ = estimated elasticity with respect to output. v is a disturbance term and the e's are the residuals from the estimated equations.

The two-tailed .05 probability level value in the t-distribution is 2.02 and the .01 level value is 2.71. Improved accuracy of our current regression program has led to trivial differences between some of the statistics in this table and corresponding statistics in Table 2 of Eisner and Nadiri (1970).

and narrowly as possible on the critical issues on which we differed.[3]

Second, of more substantive importance, Hall and Jorgenson are not correct, at least for the Jorgenson data, in their assertion that . . . "The estimated elasticity of substitution is highly sensitive to this error in specification." In response to this assertion, also included in Jorgenson and Stephenson (1969), Eisner and Nadiri (1970) have completed fairly exhaustive exploration of the parameter space in values of the first-order autoregressive parameter, ρ, with the specification of seven lagged values of each of the independent variables as in the initial Eisner-Nadiri paper, as indicated briefly in Table 1. Results confirm fully the original Eisner-Nadiri findings and leave no ground for the Hall-Jorgenson rejection either of them or, by inference, the related additional results reported by Eisner (June 1969). Maximum likelihood estimates[4] are $\hat{\rho} = .41$ and $\hat{E}_p = .0832$. For a variety of conditional values of ρ systematically covering the space between zero and one, the estimated elasticity of capital stock with respect to the preferred relative price variable, c_2, varies only trivially from the value of .0525 estimated from the first difference equation reported in the Eisner-Nadiri paper. The elasticity of capi-

[3] See Eisner and Nadiri (1968, pp. 369–70; 1970, fn. 3), and Jorgenson and Stephenson (Apr. 1967, pp. 180–81).

[4] Obtained by exploring the parameter space of ρ first in intervals of .1, between zero and unity, and then in intervals of .01 and, in case of c_1, below, of .001, in the neighborhoods of ρ in which the sum of squared residuals apparently approached a minimum.

tal stock with respect to output also varies relatively little. As noted in the pioneering work by Koyck, while distributed lag estimators using lagged values of dependent variables are not consistent if disturbances are autocorrelated, limits to the extent of bias attributable to such autocorrelation can be estimated and may in fact turn out to be narrow.

Despite our reservations about the rental price of capital, c_1, involving the earnings-price ratio, we have carried out similar estimates with it. Again, the estimated price elasticities are singularly unaffected, nowhere rising above the value of .213 (for $\rho=.9$). Maximum likelihood estimates were $\hat{\rho}=.925$, and $\hat{E}_p$, the estimated price elasticity$=.202$. The estimated elasticities of capital stock with respect to output do prove somewhat more varied when the earnings-price ratio enters into the rental price of capital but they are uniformly positive and the results are not bizarre. For the maximum likelihood estimate of $\hat{\rho}=.925$, the hypothesis of unitary elasticities with respect to both prices and output is rejected by direct t-tests (conditional on ρ) in the unconstrained regressions. And F-ratios of residuals from unconstrained regressions and regressions with elasticities constrained to be unitary, maintaining the hypotheses that errors are a first-order autoregressive process as suggested by Bischoff, indicate statistical rejection, with respect to both relative price variables, of the hypothesis of unitary elasticity of demand for capital.[5,6]

[5] See Eisner and Nadiri (1970, Table 3).

[6] Bischoff unaccountably employed a slightly different lag distribution, involving eight past quarters for each of the independent variables instead of the seven reported upon both by Jorgenson and Stephenson and by Eisner and Nadiri, but his results in reexamining the Jorgenson data prove strikingly consistent with the Eisner-Nadiri findings. Using c_2, involving the bond yield in the rental price of capital services, Bischoff reports that with his preferred, maximum likelihood estimate of the autoregressive parameter, $\rho=.5$, the estimated elasticity of substitution is .072. Turning to the data set using c_1, involving an earnings-price ratio in the rental price of capital services, Bischoff reports the maximum likelihood estimates are $\hat{\rho}=.2$ and $\hat{E}_p$, the estimated elasticity of substitution$=.358$. It is true that in this latter case Bischoff obtains an elasticity of capital with respect to output of .028, which is hardly very sensible, indicating, as Bischoff points out, that if the production function is homogeneous it is of degree -1.95, implying sharply *negative* returns to scale. But this should be at least as discomforting to Jorgenson as to me, since his model *specifies* that the elasticity of capital with respect to output is also unity. If this demonstrates anything, it is that, as Eisner and Nadiri argue, a "rental price of capital" depending upon the earnings-price ratio of equity involves an unfortunate mixture of expectational effects in the price of equity, which might well relate positively to both output and expected future output, rather than a correct measure of the cost of capital.

III. *The Bischoff Data and Putty-Clay Effects*

Hall and Jorgenson apparently claim support, from the Bischoff application of the Eisner-Nadiri model to Bischoff's own data, for their crucial assumption that a price elasticity of unity is consistent with unconstrained estimates of the investment function. It is perhaps a bit strange that Hall and Jorgenson seek such support in view of their earlier insistence that their model and the Eisner-Nadiri model " . . . are mutually contradictory and that evidence predicated on the validity of one is not relevant to the validity of the other." Be that as it may, the Bischoff data raise certain questions of their own which have been cited in Eisner (May 1969), and are further discussed in Eisner and Nadiri (1970). There are also important questions of the appropriate stochastic specification of both the Jorgenson and Eisner-Nadiri models to be considered.[7] Nevertheless, if they are to be taken seriously, the Bischoff results argue just as strongly as Eisner and Eisner-Nadiri—in some ways more so—for the inapplicability of the Hall-Jorgenson analysis to the role of tax policy in investment.

The Bischoff data deal with equipment only, not structures, and extend through 1966, while the original Jorgenson data used by Eisner-Nadiri and Eisner went only to 1962. In the Bischoff data, as shown in section A of Table 2, the price elasticity does turn out to be sensitive to specification of the autoregressive parameter, varying from .0476 for $\rho=1$, the first difference form, to .7843 for $\rho=0$, the "level" form, and maxi-

[7] See Eisner and Nadiri (1970, fn. 18).

Table 2—Maximum Likelihood Estimation of Autoregressive Parameter, Bischoff Data, 1949–66, and Bischoff Data on Equipment and Output but Jorgenson Price of Capital Variables, 1949–62, Capital Demand Equation

$$[1-\rho L]\left[\ln K_t = \gamma_0 + \sum_{i=1}^{7}\left(\gamma_{pi}\ln\left(\frac{p}{c}\right)_{t-i} + \gamma_{qi}\ln Q_{t-i}\right) + \sum_{j=1}^{2}\omega_j \ln K_{t-j} + u_t\right]$$

ρ (1)	$\sum e^2 \times 10^4$ (2)	$\hat{E}_p$ (3)	$\hat{E}_q$ (4)	$t_{(1-\hat{B}_p)}$ (5)	$t_{(1-\hat{B}_q)}$ (6)
A. Bischoff Data, 1949–66					
−0.1	1.1288	.8023	1.0372	.38	− .85
−0.01	1.124336	.7864	1.0347	.43	− .83
0.0	1.124347	.7843	1.0345	.43	− .83
0.1	1.1289	.7592	1.0334	.51	− .84
0.4	1.1859	.6544	1.0428	.80	−1.19
0.7	1.4079	.3922	1.0814	1.44	−2.10
1.0	1.7938	.0476	.8543	3.83	.73
B. Bischoff Data on Equipment and Output but Jorgenson price of capital, c_1					
−0.1	1.2568	.1957	.5300	1.38	1.93
0.0	1.2492	.1934	.5235	1.44	2.10
0.1	1.2501	.1866	.5320	1.52	2.24
0.4	1.2782	.1425	.6444	1.94	2.33
0.7	1.3272	.0972	.8131	2.58	1.54
1.0	1.5238	−.0003	.7748	2.61	.77
C. Bischoff Data on Equipment and Output but Jorgenson price of capital, c_2					
−0.1	.6497	.2912	.9905	2.74	.26
0.0	.6475	.2827	.9881	3.35	.34
0.1	.6507	.2744	.9856	3.42	.40
0.4	.6953	.2477	.9822	3.29	.43
0.7	.8107	.1952	.9991	2.93	.02
1.0	.9799	.1283	.9742	3.24	.10

mum likelihood estimates are $\hat{\rho}=-.01$ and $\hat{E}_p=.7864$. (With the Koyck-type equation involved in the Hall-Jorgenson formulations, the estimated elasticities vary from .0367 for $\rho=1$ to .6524 for $\rho=0$, with a maximum estimate of elasticity of .7153 for $\rho=.3$ and maximum likelihood values of $\hat{\rho}=.4$ and price elasticity of .6815.)

As pointed out previously (Eisner, May 1969, pp. 62–63, and Eisner and Nadiri 1970), there are some serious questions to be raised with regard to the Bischoff definition of the cost of capital entering into his rental price variable. For his cost of capital was in effect a weighted average of bond yield and dividend-price ratios where the weights were estimated from a similar body of data.[8] In an effort to isolate the cause of the divergences between the results which Bischoff reports with his own data and those that we all obtain with Jorgenson's data, Eisner and Nadiri applied Jorgenson's rental price of capital variables to Bischoff's output and capital equipment variables. The results were a striking reconfirmation of the original Eisner-Nadiri findings. Taking ρ at intervals of .1 in the range [−0.3, 1.0], the price elasticity was never estimated to be higher than .31 and the maximum likelihood estimates of price elasticities were $\hat{\rho}=0.0$ and $\hat{E}_p=.1934$ and .2827 for c_1 and c_2, respectively, as indicated in sections B and C of Table 2.

[8] As indicated in Eisner and Nadiri (1970, fn. 17), relations among the relative price variables are curious. Jorgenson's two p/c variables are negatively related to each other while Bischoff's is positively related to Jorgenson's p/c_2 but negatively related to Jorgenson's p/c_1.

Estimates of price elasticities in a model involving no lagged values of the dependent variable, and hence implying a structural equation with only seven quarterly lags for price and for output, yield no positive figures for price elasticity, regardless of the value of the autoregressive parameter, ρ. While this last might be explained by the argument, emphasized by Bischoff, that a putty-clay model is appropriate to the investment function and that responses of investment to change in relative prices are much slower than responses to changes in output, one may wonder why Hall and Jorgenson fail to discuss Bischoff's putty-clay results. For the Jorgenson model and its various applications insist that the response of investment must have the same lag distribution to each of its determinants, whether output, price of capital goods, interest rates, earnings-price ratios, tax depreciation allowances, investment tax credit, or corporate or business income tax rates. All of the Hall-Jorgenson estimates in their original paper and in their "Reply" depend just as critically on this assumption as on the one of unitary price elasticity. Indeed, if Bischoff is right in his putty-clay argument, application of the Hall-Jorgenson estimates for policy purposes would appear likely to be even more unsettling than if Eisner and Eisner-Nadiri are right in their indications of low long-run as well as low short-run effects of the tax parameters under consideration. For, as is well known, a counter-cyclical action which is potent but ill-timed is more destabilizing than one which has very little effect at any time. (See Milton Friedman.)

IV. *Conclusion*

I would hardly claim the last word on either estimates of elasticities of substitution from investment functions or the related questions of price elasticities of investment demand and effects of tax incentives. Problems of data, specification and estimation are all at best severe. But this would appear to argue all the more for facing up to persistent evidence in conflict with assumptions. To do otherwise is the path of dogma, not the sophisticated empirical analysis so much needed for informed decisions on matters of policy.

REFERENCES

C. W. Bischoff, "Hypothesis Testing and the Demand for Capital Goods," *Rev. Econ. Statist.*, Aug. 1969, *51*, 354–68.

V. K. Chetty, "Pooling of Time Series and Cross Section Data," *Econometrica*, Apr. 1968, *36*, 279–90.

R. M. Coen, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 370–79.

R. Eisner, "Investment and the Frustrations of Econometricians," *Amer. Econ. Rev.*, May 1969, *59*, 50–64.

———, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 379–88.

——— and M. I. Nadiri, "Investment Behavior and the Neo-Classical Theory," *Rev. Econ. Statist.*, Aug. 1968, *50*, 369–82.

——— and ———, "Once More on That 'Neo-Classical Theory of Investment Behavior,'" *Rev. Econ. Statist.*, May 1970, *52*, 216–22.

M. Friedman, "The Effects of a Full Employment Policy on Economic Stability: A Formal Analysis," in *Essays in Positive Economics*, Chicago 1953.

G. Fromm, ed., *Tax Incentives and Capital Spending*, The Brookings Institution, Washington, forthcoming.

R. E. Hall and D. W. Jorgenson, "Tax Policy and Investment Behavior," *Amer. Econ. Rev.*, June 1967, *57*, 391–414.

——— and ———, "Tax Policy and Investment Behavior: Reply and Further Results," *Amer. Econ. Rev.*, June 1969, *59*, 388–401.

A. C. Harberger and M. J. Bailey, ed., *The Taxation of Income from Capital*, The Brookings Institution, Washington 1969.

D. W. Jorgenson and J. A. Stephenson, "Investment Behavior in United States Manufacturing, 1947–1960," *Rev. Econ. Statist.*, Feb. 1967, *49*, 16–27.

——— and ———, "Investment Behavior in U.S. Manufacturing, 1947–1960," *Econometrica*, Apr. 1967, *35*, 169–220.

——— and ———, "Issues in the Development of the Neo-Classical Theory of Investment Behavior," *Rev. Econ. Statist.*, Aug. 1969, *51*, 346–53.

L. M. Koyck, *Distributed Lags and Investment*, Amsterdam 1954.

E. Kuh, "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," *Econometrica*, Apr. 1959, *27*, 197–214.

R. E. Lucas, "Labor-Capital Substitution in U.S. Manufacturing," in A. C. Harberger and M. J. Bailey, eds., *The Taxation of Income from Capital*, The Brookings Institution, Washington 1969, 223–24.

M. Nerlove, "Recent Empirical Studies of the CES and Related Production Functions," in M. Brown, ed., *The Theory and Empirical Analysis of Production*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth* Vol. *31*, New York 1967, 55–122.

Capital Mobility in a Tariff-Ridden International Economy

By JAMES J. RAKOWSKI*

The economic reality of international movements of capital is now gradually being incorporated into modern international trade theory. In terms of the standard two-country, two-factor, two-good model of a static international economy impeded by a tariff, investigations into the implications of capital mobility (and the resulting changes in factor endowments) have been pursued in two directions.

First, in a famous article Robert Mundell has shown that in an international economy satisfying the conditions of the factor price equalization theorem, the same equilibrium which is sustained by both factor mobility and free trade can be sustained by either free trade alone or mobility of the one factor capital alone. Thus it would seem to follow that if free trade is impeded by tariffs, capital mobility will replace trade (p. 325) and maintain undiminished the efficiency of the international economy. This result is noteworthy because it indicates a remarkable resilience in the international economy.

Second, if a country in an international economy with capital mobile wishes to maximize its own welfare at the expense of its trading partner, then it has at its disposal not only an optimal tariff on trade but also a simultaneously determined optimal tax on capital returns. Formulae have been derived by Murray Kemp (1966) and by Ronald Jones for the optimal tariff and optimal tax in a simple general-equilibrium model of an international economy.

* The author is an assistant professor at the University of Notre Dame. This article is taken from part of his Ph.D. thesis at the University of Minnesota. He is grateful to his adviser John S. Chipman, and to Anne O. Krueger, Murray Kemp, Ronald Jones, and, particularly, M. June Flanders, for help and encouragement. He also wishes to thank his colleagues Kwan S. Kim, Frank Bonello, Roger Skurski, and Jae Cho for comments at a later stage. He alone is responsible for errors.

This paper maintains that the investigations in both directions have been vitiated by an oversight regarding the full implications of the peculiar conceptualization of capital mobility upon which the results hinge. Capital mobility in a tariff-impeded economy is conceived to mean that goods which move as returns to overseas capital are unaffected by the tariffs levied on goods secured through trade. But if capital is mobile in this strong sense, then capital mobility can be shown (almost trivially) to be a substitute for trade under much weaker assumptions than previously recognized; and the taxes and tariffs purported by Kemp and Jones to be optimal are not in general even sustainable in equilibrium, for they do not satisfy a necessary equilibrium condition which we derive. And even if one allows capital to be mobile in a somewhat weaker sense, then, although it may be true that an optimal tax and tariff similar to those proposed are consistent with equilibrium, the tariff and tax of the specified type may not be all that is necessary. In cases in which capital is mobile in both directions and in which the prescribed tax is negative, another, positive tax must be added to the prescription. This paper is therefore submitted as a contribution to a more adequate conceptualization of capital mobility.

I. *The Model*

There are two countries, the home country and the foreign country. Foreign country symbols will be marked by a superscript * to distinguish them from the unmarked symbols of the home country. There are two factors of production, labor, L, and "capital," K. Two goods, good 1 and good 2, in quantities x_1 and x_2, are produced in each country (implying incomplete specialization in both countries, a strong assumption later to be relaxed) by allocating labor and capital to

two linear homogeneous production processes, which may vary between countries for each good. (Mundell's argument applies only to the special case in which the production processes are identical between countries.) Some of each factor is needed for the production of each good, and each factor can be smoothly substituted for the other. There is no joint production. The quantity of labor is given and fixed in each country. There is no saving or capital formation in either country; that is, ownership of capital is given and fixed in each country. However, any owner (capitalist) can invest his capital in either production process in either country. We thus have

$$x_1 = g(L_1, K_1) \tag{1}$$

$$x_2 = h(L_2, K_2) \tag{2}$$

$$x_1^* = g^*(L_1^*, K_1^*) \tag{3}$$

$$x_2^* = h^*(L_2^*, K_2^*), \tag{4}$$

where K_1 is that *capital of unspecified national ownership* allocated to the production of good 1 in the home country (and analogously for K_2, K_1^*, and K_2^*); but where L_1 is the *labor of the home country* allocated to the production of good 1 in the home country (and analogously for L_2, L_1^*, and L_2^*). We assume (what is equivalent to full employment of all factors) that

$$L_1 + L_2 = L \equiv \text{endowment of labor at home.} \tag{5}$$

$$L_1^* + L_2^* = L^* \equiv \text{endowment of labor abroad.} \tag{6}$$

$$K_1 + K_2 + K_1^* + K_2^* = K + K^* \equiv \text{endowment of capital in the world.} \tag{7}$$

Tastes are given in each country. Money prices need not be explicit in the model and prices determined by the system are indicated by a single ratio of exchange, p, obtained by using good 1 as numeraire. To be definite, we assume that 1) labor endowments and allocation of capital, 2) production functions, and 3) tastes are such that the home country imports good 1 if it imports anything.

There now follows a list of crucial assumptions about capital and capital mobility, taxes and tariffs. In some cases for convenience I phrase the assumptions in terms of the home-country capitalist even though an analogous assumption applies to the foreign capitalist.

We must first of all decide exactly what we mean by capital, the K in our model, so that we can justify certain assumptions about capital movements. In view of the fact that capital is an argument of the production function, that there is no saving nor capital formation, and that applicability to the above mentioned investigations is desirable, we take capital in the following sense:

(K—1) Capital is a homogeneous, durable capital good, not produced and not consumed. Capital movement is the physical movement of the capital good.

(K—2) Capital is perfectly mobile within each country.

(K—3) The actual international movement of capital is not subject to any tariffs, taxes, transport costs, or other impediments.

Note that in (K—3) we do not merely say that capital is perfectly mobile internationally. This is because we want to allow for the possibility of impediments to the flow of capital earnings, once the term is defined, and we will consider such impediments to be a diminution in capital mobility. No such impediments will arise within a country, so (K—2) can take the stronger form.

(K—4) Although capital can move between countries, the capitalist must remain in the home country and can consume only goods which are located (i.e., produced or imported) in the home country.

(K—5) Return to a unit of capital is the marginal physical product of capital in terms of that good to the production of which the unit is allocated. We write it as g_K, h_K, g_K^*, or h_K^*.

However, the capitalist can, of course, exchange some or all of this return for its equivalent value in the other good, and end up with what we call earnings. That is, more precisely:

(*K*—6) If the capital is invested in the home country, the capitalist can make the exchange at the price ratio p and then consume either good.

(*K*—7) If the capital is invested in the foreign country, the capitalist can make the exchange at the price ratio p^* and then import either good into the home country as earnings.

We now come to the assumption about capital mobility crucial to our argument. We call it the assumption of "strong capital mobility."

(*K*—8) In thus importing his earnings to the home country, the capitalist does not pay tariffs on the goods[1] which he imports.

This does seem to be how one would conceive of *perfect* capital mobility in a tariff-impeded international economy; and it is the assumption made explicitly by Mundell (p. 326 and p. 331) and Kemp (1966, p. 794, footnote 13),[2] and implicitly by Jones.[3]

(*K*—9) In importing his earnings the capitalist must, however, pay the *tax* t (some percentage of his earnings between $-\infty$ and 100), which the home country levies on goods brought home as earnings of capital.

(*K*—9*) Alternatively, if the capitalist is a foreigner taking his earnings out of the home country, he will have to pay the tax t^*, which the home country levies on the earnings of foreign capital leaving the home country.

(*K*—10) Once the capitalist gets his earnings back to the home country, he can, of course, exchange his goods at the home price ratio p.

(*K*—11) A capitalist will move his capital if (but not necessarily only if) in his reckoning by doing so he obtains disposal in the home country over more of at least one good and no less of either good. A capitalist will not move his capital if (but not necessarily only if) in his reckoning by moving it he obtains disposal in the home country over no more of either good. No capitalist takes into reckoning the effect of his action on total tax or tariff proceeds. (Thus for the purpose of *necessary* conditions for equilibrium we need not specify the distribution of proceeds.)

(*K*—12) Only the home country levies tariffs and taxes.

(*K*—13) Except for the interference of the home country's operative tariff, τ, on its import, goods markets are competitive and goods mobile. No individual takes into account the effects of his actions on tariff proceeds.

II. *Capital Mobility as a Substitute for Trade*

We first examine this model under the assumption that taxes on capital returns are zero. We show that the model is not in interior equilibrium unless prices are equalized between countries. If prices are not equalized (because of a tariff), capital will continue to move until 1) trade ceases and prices are equalized (that is, an "interior" no-trade equilibrium is reached) or 2) there is no longer any capital of the sort (specified by location and nationality) which tends to move (that is, a "corner" equilibrium is reached).

The argument to be presented is extremely simple in its intuitive content. If a tariff is levied on goods obtained through trade and not on goods obtained as capital returns, the system will tend to so allocate its capital so

[1] In our barter model the capitalist must bring home his earnings as goods. Nothing substantive is changed if we introduce money. A capitalist may then repatriate his earnings as money, but he must convert this money into goods for consumption. The equivalent assumption in this case is not that he pays no tariff on his money returns (which of course he does not), but rather that he can then use these money returns to import goods tariff-free. Introducing money into the model emphasizes the unrealism of this assumption, but does not make such an assumption inconceivable. A country might, for example, give a capitalist a certificate saying that since he earned \$100 on his capital invested abroad, he is entitled to tariff refunds on \$100 of imported goods.

[2] In some of his other work, Kemp does not make this assumption. See, for example, 1969.

[3] Jones implies this assumption both in not differing explicitly with Kemp and in writing $r=r^*$ for an equilibrium condition, where r is the return to capital in terms of the good subject to the tariff.

as to obtain the desired goods without recourse to a penalized exchange of goods between countries.

Suppose that the home country levies a positive operative tariff on its import, which to be definite we are assuming to be good 1. Then (since the denominator of p_2/p_1 becomes larger at home) we have $p^*>p$. It is evident that for the goods market to be in equilibrium we must have

$$p^* = p(1 + \tau), \tag{8}$$

where τ is the *operative* tariff, which may differ from the nominal tariff, but only if there is no trade.

But with a positive operative tariff the capital market cannot be in equilibrium. Suppose we have

$$r_1 = \overset{*}{r}_1, \tag{9}$$

where r_1 and r_1^* are defined by

$$r_1 \equiv g_K = ph_K \tag{10}$$

$$\overset{*}{r}_1 \equiv \overset{*}{g}_K = p^*\overset{*}{h}_K \tag{11}$$

Then since $p^*>p>0$, we have

$$g_K = \overset{*}{g}_K \tag{12}$$

$$h_K > \overset{*}{h}_K \tag{13}$$

The marginal physical product of capital in producing good 2 is higher at home than abroad. Capital can be shifted without cost (by K—3) from good 2 abroad to good 2 at home, where it will earn a higher return (by K—5). It follows that since foreign capatalists can repatriate their earnings free of any tariff (by K—12) and subject (by K—9*) to a tax t^*, here equal to zero, they will so shift their capital (by K—11). Therefore (9) is certainly no equilibrium condition.

It can be shown by an analogous argument that

$$r_2 = \overset{*}{r}_2, \tag{9'}$$

where r_2 and r^* are defined by

$$r_2 \equiv h_K = \frac{1}{p} g_K \tag{10'}$$

$$r_2^* \equiv \overset{*}{h}_K = \frac{1}{p^*} \overset{*}{g}_K, \tag{11'}$$

is also no equilibrium condition; for this implies

$$g_K < \overset{*}{g}_K \tag{12'}$$

$$h_K = \overset{*}{h}_K, \tag{13'}$$

and the disequilibrium indicated by (12′) leads to capital flows. Home capitalists will certainly move capital out of good 1 at home and into good 1 in the foreign country, since they can repatriate their earnings free of a tariff (by K—8) and subject (by K—9) to a tax t, here equal to zero.

It is evident that a necessary condition for equilibrium is that both

$$r_1 = r_1^* \tag{9}$$

and

$$r_2 = r_2^* \tag{9'}$$

hold; and this can happen only when $p=p^*$. Of course, the nominal tariff will still be positive, and so we must deduce that in interior equilibrium there is no trade and the tariff is inoperative, having been circumvented by capital movements. Capitalists will provide their country with the expensive good by producing it abroad, where returns are higher, and bringing home their returns.

This result does not depend upon the assumption of incomplete specialization in both countries. For example, let us assume that complete specialization prevails in both countries, with the home country producing only good 2 and the foreign country producing only good 1. A unit of capital invested in the home country earns r_2 of good 2 or (by exchange) pr_2 of good 1. A unit of capital invested in the foreign country earns r_1^* of good 1 or r_1^*/p^* of good 2. Consider the following equations:

$$pr_2 = \overset{*}{r}_1 \tag{14}$$

$$r_2 = \frac{1}{p^*} \overset{*}{r}_1, \tag{15}$$

which evidently can both hold for positive

values of r_2 and r_1^* only if $p=p^*$. Suppose that we have $p_1^*>p$. Then, for example, if (14) holds, we have

$$r_2 > \frac{1}{p^*} r_1^* \tag{16}$$

Under our assumptions this inequality leads to capital flows. A foreign capitalist can get more good 2 by investing in good 2 in the home country and bringing his returns back to the foreign country than he could get by investing in good 1 in the foreign country and exchanging his returns at the price ratio p^* for good 2. Therefore he will move his capital to the home country.

It is evident that with the flow of capital earnings completely unimpeded—that is, with zero taxes and with assumption K—8 holding—capital will tend to move in some such manner unless (14) and (15) both are satisfied. And since (14) and (15) can both be satisfied only if $p^*=p$, we see that a price differential is incompatible with interior equilibrium, and thus that our previous result, that in interior equilibrium there is no trade, holds true in the case of complete specialization also.

If an interior equilibrium exists it is identical with one in the unimpeded economy, for since prices are equalized, the tariff is inoperative. It can be removed; the economy will be unimpeded; and the equilibrium will persist.

It is proved elsewhere[4] that an interior (no-trade) equilibrium will exist if (but not necessarily only if) in unimpeded equilibrium the earnings of each country's capital endowment exceed the value of its imports.

Under this condition then we have shown that if a tariff is levied in an international economy with capital mobile in the strong sense, capital mobility will replace trade and sustain the same equilibrium formerly sustained by both trade and factor mobility. We do not need the assumptions of incomplete specialization in both countries and identical production functions, as does Mundell. Of course, we do need the unrealistic assumption of strong capital mobility; but then so also, in general,[5] does Mundell. The real impact of this demonstration is therefore to show that capital mobility is not, under realistic assumptions, a substitute for goods mobility. This is but common sense; for, after all, capital mobility entails some goods mobility.[6]

III. *Equilibrium With Tariff and Taxes*

We now show that under the assumption of strong capital mobility if taxes on capital returns and a positive tariff are levied, they must in general satisfy the condition[7]

$$(1-t)(1-t^*)(1+\tau) \leqq 1, \tag{17}$$

a necessary condition for the sustenance of equilibrium. Certainly, therefore, any combination of taxes and tariffs purported to be optimal under the assumption of strong capital mobility must satisfy (17). Even if we replace strong capital mobility with a more realistic assumption of "weak capital mobility," the condition becomes

$$(1-t)(1-t^*) \leqq 1, \tag{18}$$

and not every combination of taxes is consistent with equilibrium.

To ensure that capital is in equilibrium in our system, we must in general ensure that neither the home capitalist nor the foreign capitalist has an incentive to move capital either from the home country to the foreign country or from the foreign country to the home country.

It is convenient to divide our considerations into two cases: 1) the home country is a net debtor. 2) the home country is a net creditor. Which case prevails is of course determined by equilibrium in the specific system.

If the home country is a net debtor, we must have:

[4] See Rakowski (1968) pp. 55–64.

[5] Mundell does not need strong capital mobility if the tariff levying country is a capital importer.

[6] Labor mobility, with the laborer migrating with his services, does not entail goods mobility; and it can be proved that labor mobility is a substitute for goods mobility. The proof, however, is essentially different from Mundell's. See Rakowski (1969).

[7] If the tariff is negative, the condition is $(1-t)(1-t^*)/1+\tau \leqq 1$.

$$r_2 \geqq r_2^*(1-t) \tag{19}$$

to ensure that home capital does not flow abroad. The inequality is possible since there is no need to ensure that home capital not return from abroad, since there need be no home capital invested abroad. On the other hand, to ensure that additional foreign capital neither leaves nor enters the home country, it must hold that[8]

$$r_2(1-t^*) = r_2^* \tag{20}$$

But because prices differ between countries, it is not sufficient to consider only good 2. We must convert good 2 into good 1 at the different price ratios:

$$r_2 = r_1/p \tag{21}$$

$$r_2^* = r_1^*/p(1+\tau) \tag{22}$$

Therefore, substituting (21) and (22) into (19) and (20), it must also hold that

$$r_1 \geqq \frac{r_1^*}{1+\tau}(1-t) \tag{23}$$

$$r_1(1-t^*) = \frac{r_1^*}{1+\tau} \tag{24}$$

These conditions are certainly necessary; but, if strong capital mobility holds, (23) must be further strengthened; for if the home capitalist is not affected by the tariff, he will evidently (by K—4 and K—11) gain by moving capital to good 1 abroad as long as

$$r_1 < r_1^*(1-t) \tag{25}$$

therefore (23) must be rewritten

$$r_1 \geqq r_1^*(1-t) \tag{26}$$

The home capitalist can in effect take advantage of both price ratios.

The foreign capitalist, of course, also avoids the tariff and can take advantage of both price ratios; but it is to his advantage to make use of the higher price for good 1 in the home country, and the necessary condition remains

$$r_1(1-t^*) = \frac{r_1^*}{1+\tau} \tag{24}$$

[8] I am grateful to Murray Kemp for help in formulating these conditions.

From (24) and (26) we derive

$$(1+\tau)(1-t)(1-t^*) \leqq 1 \tag{17}$$

If the home country is a net creditor, we must have

$$r_2(1-t^*) \leqq r_2^* \tag{27}$$

and

$$r_2 = r_2^*(1-t) \tag{28}$$

or, in terms of good 1,

$$r_1(1-t^*) \leqq \frac{r_1^*}{(1+\tau)} \tag{29}$$

$$r_1 = \frac{r_1^*}{(1+\tau)}(1-t) \tag{30}$$

Once again strong capital mobility requires that one of the equations be strengthened, and this time it is (30) which must be strengthened to

$$r_1 = r_1^*(1-t) \tag{31}$$

since home capitalists need not pay the tariff; and once again we derive

$$(1-t)(1-t^*)(1+\tau) \leqq 1 \tag{17}$$

The assumption of strong capital mobility is replaced by weak capital mobility if assumption (K—8) is replaced by (K—8*):

(K—8*) In thus importing his earnings to the home country the capitalist *does* pay tariffs on the goods which he imports.

In this case the necessary condition, derived from (23) and (24) or (29) and (30), is

$$(1-t)(1-t^*) \leqq 1 \tag{18}$$

Using (18), which I consider the more realistic case and therefore the more relevant case for future research, it is helpful to illustrate graphically the possible combinations of t and t^* for the four possible situations in which the home country might find itself. This is done in Figures 1–4. The shaded areas are consistent with equilibrium.

It was only in order to be definite that these conditions were derived under the assumption of incomplete specialization in both countries. It can easily be demon-

strated that they hold for other regimes of specialization.

We conclude that any optimal taxes and tariff must satisfy these necessary conditions of equilibrium. The tariff and taxes hitherto purported to be optimal are not constrained

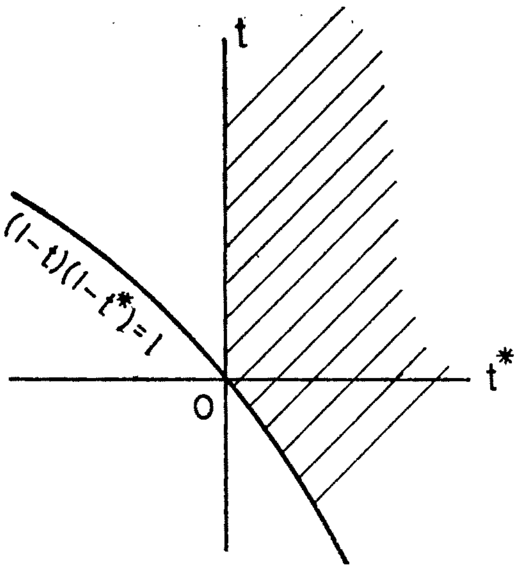


FIGURE 1.—HOME COUNTRY NET DEBTOR, $r_2 > r_2^*$ DESIRED.

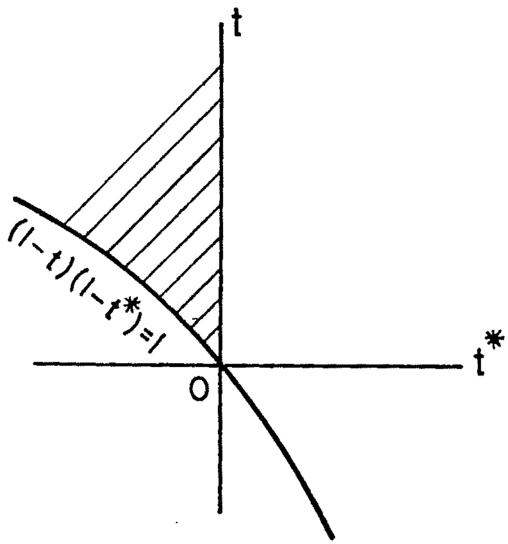


FIGURE 2.—HOME COUNTRY NET DEBTOR, $r_2 < r_2^*$ DESIRED.

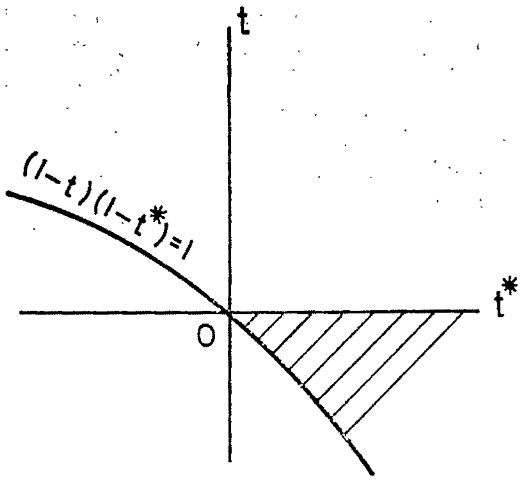


FIGURE 3.—HOME COUNTRY NET CREDITOR, $r_2 > r_2^*$ DESIRED.

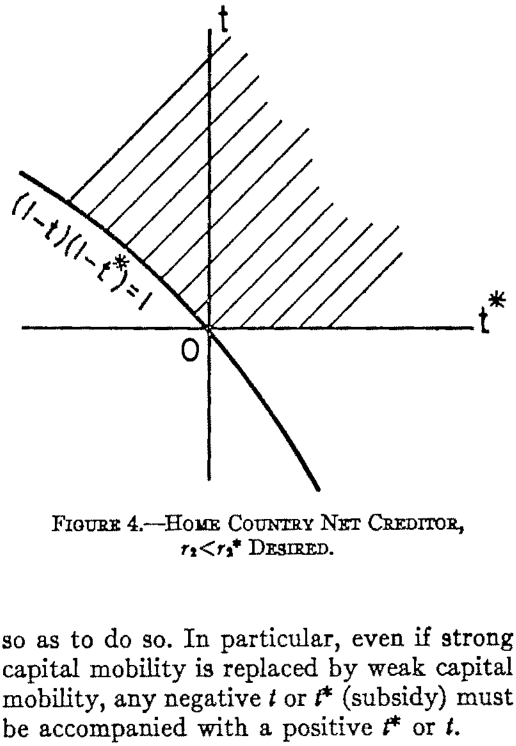


FIGURE 4.—HOME COUNTRY NET CREDITOR, $r_2 < r_2^*$ DESIRED.

so as to do so. In particular, even if strong capital mobility is replaced by weak capital mobility, any negative t or t^* (subsidy) must be accompanied with a positive t^* or t.

REFERENCES

R. W. Jones, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–38.

M. C. Kemp, "The Gains from International Trade and Investment: A Neo-Heckscher-Ohlin Approach," *Amer. Econ. Rev.*, Sept. 1966, *56*, 788–809.

———, *The Pure Theory of International Trade and Investment*, Englewood Cliffs, 1969.

R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.*, June 1957, *48*, 321–35.

J. J. Rakowski, "The Relationship Between Tariff on Goods and Taxes on Capital Returns," unpublished doctoral dissertation, Univ. Minnesota 1968.

———, "Is Labour Mobility a Substitute for Trade?" *Econ. J.*, Mar. 1969, *79*, 174–78.

Factor Mobility and International Trade: The Case of Complementarity

By Andrew Schmitz and Peter Helmberger*

By far the largest single component of *U.S.* private long-term capital outflows is direct investment.[1] During the past 10 years, *U.S.* foreign direct investment more than tripled. The average direct investment for 1953–55 was $17.7 billion which increased to an average of $54.4 billion for 1965–67.[2] In attempting to explain this phenomenal growth, it is necessary to examine how a country's foreign direct investment is related to its international product trade.

The relationship between product trade and direct foreign investment was not established in the formal theories of international trade. This is because factors of production were assumed to be internationally immobile. The well-known theories, for example, of Ricardo and Heckscher-Ohlin were concerned with explaining how product trade could arise among countries given that international labor and capital mobility were absent. In the Ricardian theory, international trade in commodities occurs because countries have different production functions, while in the Heckscher-Ohlin theory trade occurs because of international differences in relative factor endowments.

There are, however, studies available which have dealt with the question of how product trade affects international capital movements. It has been suggested that capital movements substitute for product trade; and thus, tariff barriers are largely responsible for international flows of capital. Herbert Marshall, Frank Southard, and Kenneth Taylor (p. 209), in studying *U.S.* construction of manufacturing plants in Canada, indicate that "... in the absence of tariffs the remaining barriers would be insufficient to explain the establishment of many—probably the majority—of plants now in existence." This view has been supported by Clarence Barber in a paper on "Canadian Tariff Policy." In both studies the reference is to secondary manufacturing and not to primary products and primary manufacturing. Also, their emphasis is on foreign ownership of production facilities in Canada. Thus, the international capital movements referred to are *U.S.* foreign direct investments.

In a more recent paper, Robert Mundell, within the framework of the Heckscher-Ohlin theory of trade, demonstrated that international trade in products and international capital movements are substitutes.[3] His conclusion is that product trade and capital movements are perfect substitutes, trade impediments stimulating factor movements and, likewise, increased impediments to factor movements stimulating trade. An additional implication is that, in a world of free product trade, factors do not move among countries because factor and product prices are equalized.

* The authors are assistant professor and professor, respectively, of agricultural economics, University of California, Berkeley, and University of Wisconsin, Madison. We acknowledge the comments of J. Bieri, R. Born, R. Courtney, A. E. Safarian, D. Seckler, and the reviewer of an earlier draft. However, any errors are those of the authors. Giannini Foundation paper.

[1] The distinction between direct and portfolio long-term investment (for example, foreign dollar bonds and foreign corporate stocks) is that legal control of the asset is involved in direct investment. Foreign direct investment, therefore, represents establishing a branch plant or subsidiary corporation abroad where the investor has voting control. Therefore, direct investment can be used as a measure of the *U.S.* control of corporations located in foreign countries. For a further discussion, refer to A. E. Safarian (pp. 1–2).

[2] Computed by the Department of Commerce.

[3] It should be pointed out that Mundell defined capital as a physical homogeneous factor which does not create balance-of-payments problems when it moves internationally. Also it is assumed that capitalists, as consuming units, do not move with their capital so national taste patterns are unaltered. This particular definition of capital raises certain questions which are discussed in a later section.

However, a shortcoming of the Mundell analysis is that no distinction is made among the major classifications of product trade. As indicated, Marshall, Southard, and Taylor and Barber referred explicitly to trade in secondary manufacturing when claiming that product trade and foreign direct investment are substitutes. However, a large part of international trade and direct foreign investment associated with it are in primary commodities and primary manufacturing. For example, at the time of Mundell's paper, over 60 percent of *U.S.* foreign direct investment was in primary and not in secondary manufacturing industries.

The studies available on trade in primary products and primary manufacturing would lead one to question whether or not product trade and direct foreign investment are substitutes. To a large extent, primary exports of many countries are developed by foreign capital which suggests that product trade, at least in certain primary goods, and international capital movements are complements. A striking example of this is J. V. Levin's findings for Peru in which case, despite virtually free trade, large sums of foreign capital were invested in the 19th century in Peru to develop its then major export commodity, guano. This study suggests that, in the absence of foreign capital, the Peruvian exports of guano would not have been sizable. There are, however, a whole host of world export industries which are controlled by foreign interests. Examples are Chilean copper; Bolivian tin; Canadian iron ore, potash, pulp, and paper; Australian copper; and oil production and refining facilities in Venezuela and several Near-East states, including Iran. The industries cited are largely controlled by the United States. To illustrate, the seven largest oil companies in the United States account for approximately 90 percent of the world oil production outside of North America.

It is extremely doubtful that trade in oil, copper, iron ore, etc., would have been fostered by placing impediments on the international flow of capital. To develop these industries requires substantial amounts of capital and the associated know-how, which in many countries, is simply not available from domestic sources.[4] Therefore, it appears that international capital movements and trade in primary products and primary manufacturing are not substitutes but are instead complements.[5] To support these findings, this paper demonstrates that it is theoretically possible to construct models in which long-term international investment and product trade are complements, not substitutes, in that impediments to the movement of one also impede the movement of the other.

I

In attempting to demonstrate that product trade and international capital movements can be complements rather than substitutes, as previous authors have argued, it appears necessary to relax at least one of the assumptions underlying the Heckscher-Ohlin theory apart from the assumption of international factor immobility. That is the assumption of identical international production functions. As mentioned, Mundell's demonstration that international capital movements and product trade are substitutes is based on the Heckscher-Ohlin framework which assumes, among other things, identical production functions and similar tastes; for, if production conditions are identical, it is not immediately obvious why capital would be invested abroad to develop primary industries unless perhaps because of lower input costs. In addition, although not as crucial as the production function assumption, it will become apparent that different demand conditions should be assumed to exist among regions.

[4] For an excellent discussion, see Robert Baldwin (1963) and Anne Krueger.

[5] Whether or not *U.S.* foreign direct investment and product trade in secondary manufacturing are substitutes is even debatable. See, for example, the empirical studies by Anthony Scaperlanda and Ralph d'Arge. Also, in this connection, it should be pointed out that foreign direct investment, in certain cases, is not a function of the degree of tariff protection. That is, foreign direct investment can occur even in a world of free trade in secondary manufacturing which, as Mundell demonstrated, is ruled out when using the Heckscher-Ohlin framework of analysis. Some individual studies which suggest this to be true are those by Aharoni, John Dunning, and Dudley Phelps.

The reason for assuming different production functions among countries when dealing with primary commodities should be apparent.[6] For primary products, such as copper, oil, timber, and potash, production conditions are not identical among regions of the world. In certain countries some of these natural resources do not exist. Among regions where they do exist, the relative costs of developing these resources differ widely. One cause is the poor ore quality and the depth of mining required in some areas relative to those with large natural supplies.[7]

Like production, different total demand conditions also exist among countries. For most primary products, the demand, after population and income differences are taken into account, is larger for the country supplying capital than for the recipient country. For example, as pointed out earlier, the United States has a large market for oil as compared to either Canada, Iran, or Venezuela. Likewise, there are virtually no local markets for potash mined in Canada. Therefore, any sale of Canadian potash must be to foreign markets.[8]

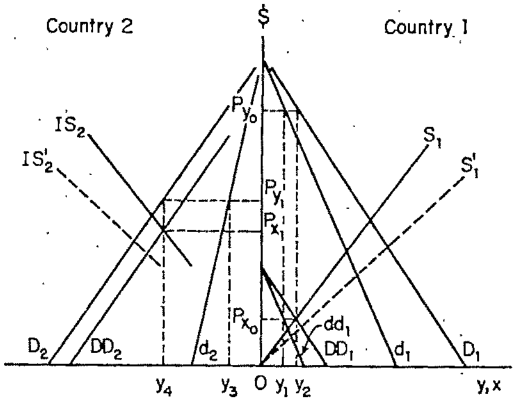


FIGURE 1. A SPATIAL MODEL OF TRADE EXCLUDING AND INCLUDING FACTOR MOBILITY.

To illustrate that product trade and international capital movements can be complements, a spatial equilibrium framework is used. The analysis is shown graphically in Figure 1. It is assumed that, in the absence of international factor movements, capital, which is denoted by x, is available only in country 1.[9] The supply curve of capital in country 1 is OS_1 while the domestic demand for product y is d_1. In country 2, d_2 represents the domestic demand for y. To be consistent with the previous discussion, d_1, which corresponds to the country supplying the capital x, is substantially larger than d_2. In the model, transportation costs are excluded.

Assuming that factors of production are internationally immobile, production will only occur in country 1 since x is not available in the other country. Therefore, country 1 produces product y for both markets; thus, the relevant product demand curve is the aggregate of d_1 and d_2 which is represented by D_1. Equilibrium production is oy_2, which is the amount determined by the intersection of the supply curve of x (not shown) and DD_1 which is the aggregate derived demand for

[6] Studies which have refuted the notion of identical international production functions include those by Kenneth Arrow, Hollis Chenery, Bagicha Minhas, and Robert Solow. Also, see the study by Kreinin.

[7] It is our contention that the larger the differences in production conditions among countries in primary products the greater will be the flow of international capital.

[8] How the size of a domestic market relative to foreign markets affects investment decisions remains an empirical question. One would expect, however, that the smaller the domestic market relative to foreign markets the more reluctant are domestic investors to develop a given product, especially one which requires large initial sums of capital. In addition, even though a country has a large natural supply of such commodities as oil, tin, copper, and potash, if the potential domestic demand for these is small, then potential demand in foreign markets may go unnoticed by domestic investors. These—both the high cost of information and the unavailability of information—appear to be among the major reasons why the United States, France, and West Germany invested large sums of capital in Canada to mine its potash. Raymond Vernon (p. 202), in discussing product specialization for less-developed countries, states:

> In projecting the patterns of exports from less-developed areas, however, we cannot afford to disregard the fact that information comes at a cost, and that entrepreneurs are not readily disposed to pay the price of investigating overseas markets of unknown dimensions and unknown promise. Neither are they eager to venture into situations which they know will demand a constant flow of reliable marketing information from remote sources.

[9] The question as to the meaning of capital is raised in the last section.

x.[10] The latter corresponds to D_1 and the equilibrium product price P_{y_0}. In the model, both countries pay P_{y_0} for product y, and factor x receives a payment P_{x_0}. Country 1 exports y_1y_2 of product y to country 2.

The next step is to demonstrate the case in which not only products can be traded but also factors can move between countries. To be consistent with our earlier discussion, production functions are assumed to be different between regions, and x is much more productive in country 2 than in country 1. As noted earlier, x is unavailable to country 2 from domestic sources. This being the case, it is necessary to plot the excess supply curve of x in the second quadrant in Figure 1. This is represented by IS_2. Given the aggregate demand curve of d_1 and d_2 represented as D_2, if all of y is produced in country 2, it is possible to determine equilibrium trade flows of factor x and product y and equilibrium prices for both. The equilibrium is represented by the intersection of IS_2 and DD_2 which is the aggregate derived demand curve for x.[11] The corresponding equilibrium product price is P_{y_1}; for x, it is P_{x_1}. In this situation, oy_4 of product y is produced of which y_3y_4 is exported to country 1. Therefore, by comparing this analysis with that earlier, which allows for only trade in the product, trade increases when factors are allowed to move internationally. Therefore, using free trade in both the factor and product as a norm, factor and product movements are not substitutes—they are complements. This is because in the model country 2 is much more productive than country 1 in the production of the product. Consequently, by allowing factor mobility, the product has become so much cheaper that total use has expanded—and, with it, the volume of trade.

In Figure 1, the aggregate derived demand curve DD_2 was arbitrarily drawn such that in equilibrium none of y is produced in country 1. How the derived demand curve for x is drawn depends on the production function specified for each region.[12] The analysis demonstrates that the greater the differences in international production functions the greater will be the flow of international capital from the capital-intensive country. This is clearly indicated since, if the production function in country 2 becomes similar to that in country 1, due, for example, to a change in technology, DD_2 would shift to the right. This would decrease product trade, increase product prices, decrease capital flows, and decrease factor prices. In addition, the volume of factor movements depends on the responsiveness of the factor to changing factor prices. Consider a new supply equation for x represented by S_1' and the corresponding excess supply IS_2'. It can be demonstrated, with reference to the derived demand equation for x, that the new equilibrium price is below P_{y_1}; thus, the volume of product trade increases as does the transfer of x between regions. This is so since the shift of DD_2 to the right is not of sufficient magnitude to intersect IS_2' directly below IS_2 and DD_2. Therefore, with reference to the relationship between capital movements and trade in primary products and primary manufacturing, the model suggests that volume of international capital movements depends on the response of capital to changes in interest payments and the extent to which there exist international differences in production functions.

II

To this point considerable emphasis has been placed on distinguishing among types

[10] Actually, when deriving the demand for x corresponding to the equilibrium value for y and P_y, one merely determines one quantity for x. P_x is obtained from knowing the supply curve of x. But for completeness and ease of illustration, rather than merely representing a point on the supply curve of x corresponding to an equilibrium situation, the derived demand curve is drawn although it need not be drawn parallel to D_1. In addition, DD_1 is drawn to correspond with the equilibrium product price P_y and is, therefore, not the derived demand curve for alternative product prices corresponding to the entire D_1 curve.

[11] Again, there is no specific reason that DD_2 has to be drawn parallel. Also, as in the earlier example of allowing only trade in the product, DD_2 is drawn only with reference to a given price of y which in equilibrium is given by P_{y_1}. DD_2 does not represent the derived demand for x for all product prices corresponding to D_2.

[12] In Figure 1, different production functions are represented by a small vertical distance between D_2 and DD_2 relative to that between DD_1 and D_1.

of products which are traded internationally. For empirical purposes, however, it is also necessary to delineate the types of international capital movements which are affected by imposing restrictions on the flow of products among countries.

Carl Iverson and George Borts, for example, in their writings on long-run international movements, conceive a capital movement to be a transfer of goods and services between countries offset by an opposite movement of private titles to ownership of private-debt instruments. Thus, a country which is importing capital has a surplus in its balance on capital account and a deficit in its balance on current account; that is, the level of investment in the country exceeds the level of savings. As pointed out at the outset, if one considers only the largest component of long-term private *U.S.* capital movements—direct investment—only those assets abroad have to be considered which are legally controlled by *U.S.* firms. This represents establishing a branch plant or subsidiary corporation abroad where the investor has voting control.

Recently, Baldwin and H. Oniki and Hirofumi Uzawa have demonstrated models of international trade which explicitly incorporate trade in capital goods. These, however, are capital goods exports, such as machinery and equipment, which are paid for with goods produced in the country importing the physical capital. Therefore, this type of analysis is consistent with the Heckscher-Ohlin framework, and it does not imply that factors of production are internationally mobile. The above models therefore incorporate trade in capital goods along with the assumption that capital is internationally immobile.[13] This is because of the distinction made between capital goods trade, which is the analysis by Baldwin and Oniki and Uzawa, and international long-term capital movements. In the latter case, a capital export is defined in equilibrium to be the positive excess of savings over investment, equal in magnitude to the current account balance of payments. Therefore, capital goods exports do not imply an export or import of capital.

At the outset, it was pointed out that the studies by Marshall, Southard, and Taylor and Barber used the concept of private *U.S.* direct investment and how it was affected by restricting product trade. However, this does not appear to be the concept of capital used in Mundell's demonstration (p. 322) that product trade and international capital movements are substitutes since capital is defined as a physical, homogeneous factor which does not create any blance-of-payments problems when it moves internationally. This implies a concept of capital similar to capital goods trade rather than capital exports of a long-run nature.[14]

It is our contention, however, that the relationship between product trade and capital goods trade is not the same as between product trade and long-term international capital movements. Thus, it is necessary to distinguish among different forms of capital movements. Restrictions placed on international trade in products may have a substantial impact on *U.S.* foreign direct investment with little or no effect on *U.S.* capital goods trade. Consider, for example, the recent large increase in *U.S.* direct investment in Europe which many argue has resulted from the common external tariff established by the European Economic Community. Part of this increase in investment has been a result of mergers between *U.S.* and European firms.[15] These mergers

[13] As Baldwin (p. 841) points out: "It is perhaps not out of place to remind ourselves that the traditional assumption of trade theory concerning the immobility of capital refers to abstract capital and not capital goods."

[14] Since, in the Heckscher-Ohlin theory, the assumption of international capital immobility does not refer to capital goods, it could be argued that Mundell never really relaxed the assumption of international capital immobility.

[15] Although data are scanty, they suggest that there has been a phenomenal increase in international merger activity between United States and European firms. In 1968, for example, Monadnock Paper Mills, Inc., of Bennington, New Hampshire merged and now has control of the leading paper manufacturer in Austria—Lenzenger Zellulose and Papierfabrik,A.G. Likewise, Scott Paper Company merged with Bouton Brochard of France and has voting control in the so-called international firm. One last example is Litton Industries which

generally involve international transfers of technical entrepreneurial skills accompanied by little or no flows of physical capital. However, even in cases where mergers are not involved, *U.S.* industrial complexes in Europe can purchase, from internal sources, most of the physical capital needed, thus eliminating the need for capital goods exports from the United States. In addition, it should be pointed out that *U.S.* direct foreign investment does not necessarily even involve large transfers of liquid capital from the United States. J. Servan-Schreiber (p. 14) points out that although in 1965 the United States invested $4 billion in Europe only 10 percent of this was a direct dollar transfer from the United States. This is due to: 1) American investors obtaining loans from the European capital market to finance investments abroad; 2) direct credits from European countries and government subsidies; and 3) internal financing from local earnings. Therefore, the major component of the recent phenomenal growth in *U.S.* foreign direct investment appears to be international movements of human capital and not capital goods exports or direct dollar transfers from the United States.[16]

III.

In this paper we have demonstrated that it is possible to construct models in which factor and product flows can be substitutes or complements depending on the framework of analysis and assumptions used. We have demonstrated that in at least one rather common economic situation—that is, when capital inputs are exported because natural resources and primary manufactured products can be imported—factor and product flows are complements.

The extent to which product trade and international capital movements are substitutes or complements is an empirical question and is not dealt with here. We merely emphasize the points raised earlier. It was demonstrated that, in order to carry out quantitative analysis to determine how product trade affects international capital movements, both the type of commodity traded and the form of international capital transactions must be specified. It appears that the relationship between international capital movements and exports of primary products and primary manufacturing is not the same as between capital movements and exports of secondary manufacturing. Also, the relationship between product trade and capital goods exports appears to be different from that between product trade and long-term international capital movements. Hopefully, future empirical studies will determine how restrictions on different types of product trade affect different forms of international capital flows.

acquired control of Grundig Organization of Triumph and Adler factories at Frankfurt and Nuremberg, Germany.

[16] In our demonstration in the previous section, the capital input was not defined. If defined as physical capital, as in the Mundell analysis, the analysis would demonstrate the complementarity relationship between product trade and capital goods exports. However, to demonstrate complementarity between product trade and foreign direct investment—a specific type of long-run international capital movements—the definition of capital must include technical and entrepreneurial know-how which, as pointed out, is a crucial component of *U.S.* foreign direct investment.

REFERENCES

Y. Aharoni, *The Foreign Investment Decision Process*, Cambridge 1966.

K. J. Arrow, "The Economic Implications of Learning by Doing," *Rev. Econ. Stud.*, June 1962, *29*, 155–73.

———, H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *53*, 225–50.

R. E. Baldwin, "Export Technology and Development from a Subsistence Level," *Econ. J.*, Mar. 1963, *73*, 80–92.

———, "The Role of Capital-Goods Trade in the Theory of International Trade," *Amer. Econ. Rev.*, Sept. 1966, *56*, 841–48.

C. L. Barber, "Canadian Tariff Policy," *Can. J. Econ.*, Nov. 1955, *21*, 513–30.

G. H. Borts, "A Theory of Long-Run International Capital Movements," *J. Polit. Econ.*, Aug. 1964, *72*, 341–59.

R. d'Arge, "Note on Customs Unions and Direct Foreign Investment," *Econ. J.*, forthcoming.

J. H. Dunning, *American Investment in British Manufacturing Industry,* London 1958.
C. Iverson, *Aspects of the Theory of International Capital Movements,* London 1936.
M. E. Kreinin, "Freedom of Trade and Capital Movement—Some Empirical Evidence," *Econ. J.*, Dec. 1965, *75,* 748–58.
A. O. Krueger, "Factor Endowments and Per Capita Income Differences Among Countries," *Econ. J.*, Sept. 1968, *78,* 641–59.
J. V. Levin, *The Export Economies,* Cambridge, Mass. 1960.
H. Marshall, F. A. Southard, and K. W. Taylor, *Canadian-American Industry: A Study in International Investment,* New Haven 1936.
R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.,* June 1957, *47,* 321–35.
H. Oniki and H. Uzawa, "Patterns of Trade and Investment on a Dynamic Model of International Trade," *Rev. Econ. Stud.*, Jan. 1965, *32,* 15–38.
D. M. Phelps, *Migration of Industry to South America,* New York 1963.
A. E. Safarian, *Foreign Ownership of Canadian Industry,* Toronto 1966.
A. Scaperlanda, "The EEC and U.S. Foreign Investment: Some Empirical Evidence," *Econ. J.*, Mar. 1967, *77,* 22–26.
J. J. Servan-Schreiber, *The American Challenge,* New York 1968.
R. Vernon, "International Investment and International Trade in the Product Cycle," *Quart. J. Econ.*, May 1966, *80,* 190–207.
U.S. Department of Commerce, Office of Business Economics, *Survey of Current Business,* Washington, various issues.

A Note on Equality of Educational Opportunity

By LEWIS C. SOLMON*

"It will be profitable as mere investment to give the masses of the people much greater [educational] opportunities than they can generally avail themselves of" . . . "General ability depends largely on the surroundings of childhood and youth."

Marshall-*Principles*

Most studies of equality of educational opportunity have begun by looking at the roles played by sociological factors and by various inputs of school systems in achieving the school's goals. Goals include increasing the incomes of students in their postschool lives or, in the shorter run, maximizing Scholastic Aptitude Test scores of students (see Samuel Bowles). Finis Welch concludes that the systematic differences in returns to schooling by states are determined by the multiple of quality of schooling and the value of the marginal product of education. James Morgan and Ismail Sirageldin report evidence of association between the amount a state spends per year on primary and secondary education and the subsequent earnings of people who went through those school systems. Some might conclude from the previous studies that the way to effect "equality of opportunity" in education for subgroups in society is to increase expenditures per student by raising teacher salaries and reducing classroom size in schools attended by these underprivileged groups (see James Coleman et al.,). Others might even use the significance of the race variable as an argument (incorrectly in my view) for busing students to reduce the proportion of blacks in the ghetto schools.

It is argued here that although higher salaries and smaller classes, in ghetto schools as compared to suburban schools, will move toward equating quality of schooling available for those who attend, equality of educational opportunity will not be achieved because of inability to act on an available opportunity. To achieve equality of educational opportunity for low income students, we must take cognizance of the different effects family incomes and opportunity costs (foregone earnings of students) have on decisions about schooling by families of different socioeconomic classes. Family income is important because its level affects the ability of the family to provide at home complements to in-school learning. The importance of opportunity costs arises because any amount (or percentage) of family income which must be foregone when a potential earner stays in school probably will hurt a poor family more (or in different ways) than will the same amount foregone by a wealthy family.

The latter point is based on the assumption of diminishing marginal utility of income. More precisely, it is asserted that the sacrifice of any *percentage* of family income becomes more "painful" as family income declines. In the present context, adjustments necessitated by a $300 reduction in income for a family earning $3,000 probably hurt a youth's ability to benefit from schooling more than does a $600 reduction (or even a $700 reduction) in income of a family earning $6,000.[1]

The calculations made in Table 1 show that a larger portion of family income must be foregone, on average, when a nonwhite or Southern family sends a child to school, compared to other families. If it is assumed that a family sending its teenager to school gives up the median income of 14 to 19 year olds, then a Southern nonwhite family sacrifices 26.56 percent of its median family income by so doing. The typical family in the

* City University of New York and the National Bureau of Economic Research.

[1] This argument is analogous to the argument for progressive taxation which asserts that in order to equalize the "burden" of the tax, higher income groups should be taxed at *higher* rates rather than at the same rate as all groups.

TABLE 1—FRACTION OF FAMILY INCOME CONTRIBUTED BY A 14 TO 19 YEAR OLD CHILD

	U.S. Total	U.S. Nonwhite	North-east	North Central	South Total	South Nonwhite	West
(1) Median family income	$5,663	3,169	6,195	5,898	4,467	2,319	6,342
(2) Income of those aged 14–19	$ 700	650	713	708	670	617	750
(3) Fraction of family income contributed by a child aged 14–19	.1236	.2051	.1150	.1198	.1564	.2656	.1181
(4) Enrollment per population aged 14–19	.7394	.6921	.7488	.7560	.7076	.6949	.7650
(5) Fraction of pupils enrolled attending daily	.900	.889	.898	.888	.892	.881	.929
(6) Fraction of those aged 14–19 attending daily	.6654	.6152	.6724	.6713	.6311	.6122	.7106

Sources

(1) *1960 Census*, Subject Report, Sources and Structure of Family Income, Table 2.
(2) *1960 Census*, Subject Report, Educational Attainment, Table 6. (Weighted by total population aged 14–15, 16–17, 18–19, in each group.)
(3) (2) divided by (1).
(4) *1960 Census*, Subject Report, School Enrollment, Table 2.
(5) *Statistics of the State School Systems 1959–1960*, p. 46. (Nonwhite is the regional total figure multiplied by the ratio of the South to the *U.S.* total.)
(6) (5) multiplied by (4).

United States as a whole gives up only 12.36 percent of its income by sending a comparable child to school. It is interesting to note that days attended per population aged 14 to 19 varies inversely with the percent of family income foregone by sending a child to school and varies directly with median family income.

These calculations are necessarily crude. However, the point to be made is still valid if the fraction of family income foregone were equal for different groups as long as the levels of incomes of the two groups are different. Table 2 attempts to refine the estimates of earnings foregone by replacing median income of those 14 to 19 with the difference between earnings of youths who were in school and youths not in school. The data here show almost identical shares of family income foregone in the four race-region classes. Of course, if students work during out of school hours they reduce the burden of lower family income but also reduce out of school time available to be spent on their studies.

Three completely different problems must be dealt with in order to bring about equality of opportunity in education. The usual point is that to improve quality of ghetto schools, the expenditures per student in these schools must be increased. However, the differential impact of foregone earnings points to a second problem. In order to get the poor to send their high school aged children to schools of *any* quality, the cost in terms of proportionate reduction in family income must be reduced for low income families. Perhaps direct payments to these families based on sending their children to school (rather than welfare payments based on the *number* of children) are the answer. This policy would reduce the immediate cost of the decision to enroll the child in school. This point may be minimized by claims that poor families can lessen the income loss by resorting to capital markets. However, capital markets are imperfect, with respect both to availability and cost of funds to the poor, and to the traditional dislike of many families to borrow at all.

It has been suggested that since the poor have lower opportunity costs than the rich, they should demand more schooling. That is, if both groups have the same demand curve relating opportunity costs (the price to the individual) and quantity of schooling, the price is lower for the poor and so, quantity should be larger. However, both lower

TABLE 2—EARNINGS FOREGONE BY TEENAGE STUDENTS AS A FRACTION OF FAMILY INCOME

	North Whites	North Nonwhites	South Whites	South Nonwhites
(1) 1959 earnings for youths not in school in 1960	761.6	466.1	579.0	292.4
(2) 1959 earnings for youths in school in 1960	209.9	107.7	173.8	96.9
(3) Earnings foregone	551.7	358.4	405.2	195.5
(4) Median family income	6,306	4,191	4,913	2,319
(5) Fraction of family income foregone by sending a child to school	8.75	8.55	8.25	8.43

Source:

(1) and (2) from unpublished tables of Giora Hanoch. Youths are 14 to 18 years.
(3) (1) minus (2).
(4) See Table 1, source 1. Incomes for North whites and nonwhites and Southern whites were obtained by taking weighted averages using data on number of families in each group.
(5) (3) divided by (4).

income and lower rates of return are factors causing the demand curve of the poor to be lower than the demand curve of the rich. It cannot be argued that poor families should have a "longer run view" of the education decision and look at rates of return rather than merely at costs. Gary Becker (1964) has shown that rates of return are lower for nonwhites (usually the poor class) than for whites. Probably the poorer the family, the shorter its time horizon for decision making.

The differential impact of opportunity costs implies that children from poorer familes get lower quality education (or less opportunity to obtain equal quality education) even if they attend the *same* school as children of wealthy families. The loss of 10 percent of income for a family earning $3,000 has several implications: certain goods and services formerly purchased in the market will have to be produced in the home. According to Becker's theory of allocation of time (1965), as income declines when a family sends a child to school, household time is substituted for market goods in producing consumer utility. Probably, the student will have to reduce his nonschool study time in order to help with the household activities. Moreover, other members of the family will have less time to spend in assisting and encouraging the student. This lack of home stimulation has been demonstrated to be a problem in low income families which has lowered achievement in school of children from these families. Consumption of certain commodities will have to be reduced. These may include books, lamps, clothing, and nourishing foods, all complementary to schooling.

Pressure to take part-time jobs and increase household production, lower levels of complementary goods, and a less favorable home environment reduce the student's ability to get the most out of his school attendance. Compared to the student from the home having a high median income the poor student's opportunity to benefit from education has been reduced.

Finally, it has been suggested that since on average, poor black families have more children than wealthy white families have, the income sacrificed by sending any *one* child to school is less for the poor. Table 3 provides information on this point. Rather than using median regional incomes as in Tables 1 and 2, take median income for a nonwhite family with three or more earners ($5,357) and median income for all families with two wage earners ($6,671). This adjusts incomes to allow for more people of working age in a black family. If these are divided into earnings of the corresponding

TABLE 3—MEDIAN INCOME BY FAMILY SIZE AND NUMBER OF EARNERS

	U.S. Total	White	Non-White
Median Family Income When Number of Persons in Family is:			
2	4,651		
3	5,861		
4	6,289		
5	6,309		
6	6,093		
7+	5,349		
Median Family Income When Number of Wage Earners is:			
0	1,499	1,577	973
1	5,179	5,379	2,757
2	6,671	6,911	4,099
3+	8,775	9,152	5,357
Population per Household	3.29	3.23	3.85
Average Family Size	3.65	3.58	4.31

Source: Census of Population 1960.

14 to 19 year old, the black family appears to sacrifice 12.13 percent of its income by sending one child to school whereas the *U.S.* average sacrifice is 10.49 percent. Although the difference has narrowed, the black family still gives up a higher percentage of family income. Moreover, as noted earlier, an assumption of diminishing marginal utility of any fraction of income implies that even equal percentage sacrifices will hurt a poor family more than a rich family.

Conclusions

Certainly attempts to buy higher quality teachers and to have them teach smaller groups will increase the educational opportunities of students from poor families. However, there are definite limits to the effectiveness of this policy. The cost in terms of proportionate reduction in opportunity cost must be reduced for poor families to move toward real equality. Children from poor families who attend school reduce family income by a larger percentage than do children from rich families. Hence, the poor family is likely to consider school a greater hardship and so is less likely to enroll its children. Moreover, any percentage reduction in family income (due to school attendance) is more likely to cause adjustments in living patterns that will hurt a student's ability to benefit from schooling if the family is poorer.

Incentive payments for poor families to send their children to school would reduce the differential influence of opportunity cost on the decision to enroll, would enable the greater availability of household goods which complement schooling, and would increase the rate of return to schooling for the children of poor families.

REFERENCES

G. S. Becker, *Human Capital, A Theoretical and Empirical Analysis with Special Reference to Education,* Princeton 1964.

———, "A Theory of the Allocation of Time," *Econ. J.,* Sept. 1965, *75,* 493–517.

S. Bowles, "Towards an Educational Production Function," paper presented at the Conference of the National Bureau of Economic Research on "Education and Income" in Madison, Wisconsin, Nov. 15–16, 1968.

J. Coleman et al., *Equality of Educational Opportunity,* Washington 1966.

G. Hanoch, "Personal Earnings and Investment in Schooling," unpublished doctoral dissertation, Univ. Chicago, 1965. Also unpublished tables by Hanoch.

J. Morgan and I. Serageldin, "A Note on the Quality Dimension in Education," *J. Polit. Econ.,* Sept./Oct. 1968, *76,* 1069–77.

L. Solmon, "Capital Formation By Expenditures on Formal Education, 1880 and 1890," unpublished doctoral dissertation, Univ. Chicago, 1968. Especially Appendix E, "Capital Formation by Education, 1960."

F. Welch, "Measurement of the Quality of Schooling," *Amer. Econ. Rev.,* May 1966, *56,* 379–92.

NOTES

EIGHTY-THIRD ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

Cobo Hall, Detroit, Michigan, December 27–30, 1970

Preliminary Announcement of the Program

Sunday, December 27, 1970

2:00 P.M. EXECUTIVE COMMITTEE MEETING

Monday, December 28, 1970

8:30 A.M. ROUND TABLE SESSION ON THE ACADEMIC LABOR MARKET[1]
Chairman: W. LEE HANSEN, University of Wisconsin
Papers: LINDSEY R. HARMON, National Research Council
Prospective Supply in the 70's
FRANCIS BODDY, University of Minnesota
Recent Behavior of Economists' Salaries
DAVE BROWN, Miami University
Improving the Operation of the Labor Market
ROBERT STRAUSS, University of North Carolina
The Younger Economists' View of the Market
ALLAN CARTTER, New York University
Whither the Market?

8:30 A.M. TELEVISION: OLD THEORIES, CURRENT FACTS, AND FUTURE POLICIES[1]
Chairman: PETER O. STEINER, University of Michigan
Papers: R. EDWARD PARK, The RAND Corporation
The Growth of Cable TV, and Its Probable Impact on Over-the-Air Broadcasting
JOHN PETERMAN, University of Chicago
Television Rate Structures
HARVEY J. LEVIN, Hofstra University
Program Duplication, Diversity, and Effective Viewer Choice
EDWARD GREENBERG AND HAROLD BARNETT, Washington University
Program Diversity—New Evidence and Old Theories
Discussion: MERTON J. PECK AND JOHN MCGOWAN, Yale University

8:30 A.M. THE RISING COST OF HEALTH CARE: CAUSES AND POTENTIAL REMEDIES (Joint Session with The Health Economics Research Organization)
Chairman: DONALD E. YETT, University of Southern California
Papers: DAVID S. SALKEVER, National Center for Health Services, U.S. Public Health Service
A Micro-Econometric Study of Hospital Cost Inflation
VICTOR R. FUCHS and MARCIA J. KRAMER, National Bureau of Economic Research
Expenditures for Physicians' Services in the U.S.
HERBERT E. KLARMAN, New York University
Policy Alternatives for Controlling Health Service Expenditures
Discussants: RALPH ANDREANO, University of Wisconsin
SYLVESTER E. BERKI, University of Michigan
MICHAEL INTRILIGATOR, University of California, Los Angeles and University of Southern California
DONALD E. YETT, University of Southern California

[1] To be published in the *American Economic Review Proceedings*, May 1971.
[2] To be published in the *American Economic Review*, March 1971.
[3] To be published by the American Statistical Association, Business and Economics *Proceedings*.
[4] To be published in the *History of Political Economy*.
[5] To be published in *The American Economist*.
[6] To be published in the *Journal of Finance*.
[7] To be published by The National Bureau of Economic Research.

8:30 A.M. ISSUES IN GRANTS ECONOMICS (Joint Session with the Association for the Study of The Grants Economy)
Chairman: KENNETH E. BOULDING, University of Colorado
Papers: MARTIN PFAFF and ANITA B. PFAFF, Wayne State University
How Equitable are Implicit Public Grants? The Case of the Individual Income Tax
HENRY AARON, University of Maryland and Brookings Institution, and GEORGE VON FURSTENBERG, Indiana University
How Inefficient are Transfers in Kind? The Special Reference to Housing
DAVID O. PORTER, University of California, Riverside, and DAVID C. WARNER, Wayne State University
How Effective are Grantor Controls? The Case of Federal Aid to Education
MURRAY L. WEIDENBAUM, U.S. Treasury and Washington University
How Rational is Public Sector Decisionmaking? An Examination of the Norms Determining Federal Priorities

10:30 A.M. URBAN GROWTH AND DEVELOPMENT[1]
Chairman: JOHN F. KAIN, Harvard University
Papers: BRYAN ELLICKSON, University of California, Los Angeles
Jurisdictional Fragmentation and Residential Choices
EDWIN S. MILLS, Princeton University
Market Choices and Optimum City Size
HAROLD M. HOCHMAN and C. WORTH BATEMAN, The Urban Institute
Social Problems and the Urban Crisis: Does Public Policy Make a Difference?
Discussants: ROLAND ARTLE, University of California, Berkeley
STEPHEN P. DRESCH, National Bureau of Economic Research
ANTHONY H. PASCAL, The RAND Corporation

10:30 A.M. THE ALLOCATION OF SOCIAL RISK[1] (Joint Session with The Econometric Society)
Chairman: JACOB MARSCHAK, University of California, Los Angeles
Papers: WILLIAM BRAINARD, Yale University and F. TRENERY DOLBEAR, JR., Brandeis University
Financial Markets and Social Risk
MARK PAULY and RICHARD KIHLSTROM, Northwestern University
Insurance and the Optimal Allocation of Risk
RICHARD ZECKHAUSER, Harvard University
Second Best with Incomplete Contingent Claims Markets
Discussants: MARTIN BAILEY, University of Rochester
HYMAN MINSKY, Washington University
DONALD HESTER, University of Wisconsin

10:30 A.M. STATISTICAL APPRAISALS OF RECENT SHORT AND LONG TERM MONETARY AND FISCAL POLICY[1] (Joint Session with The American Statistical Association)
Chairman: THOMAS MAYER, University of California, Davis
Papers: THOMAS F. CARGILL, Purdue University
Federal Open Market Operations: A Long- and Short-Run View
VITTORIO BONOMO and CHARLES SCHOTTA, Virginia Polytechnic Institute
The Efficacy of Recent Federal Reserve Monetary Policies
KEITH M. CARLSON, Federal Reserve Bank of St. Louis
A Monetarist Model for Stabilization Policy
Discussants: DAVID FAND, Wayne State University
WILLIAM YOHE, Duke University
MARK WILLES, Federal Reserve Bank of Philadelphia

2:30 P.M. POPULATION AND ENVIRONMENT IN THE UNITED STATES[1]
Chairman: JOSEPH J. SPENGLER, Duke University
Papers: JOSEPH L. FISHER, Resources for the Future
Impact of Population on Environment
RICHARD EASTERLIN, University of Pennsylvania
Does Population Adjust to the Environment?
GLEN CAIN, University of Wisconsin
A Conventional Economist's View of the Population Bomb
Discussants: T. PAUL SCHULTZ, The RAND Corporation
PAUL DEMENY, The East-West Center, University of Hawaii
ROGER REVELLE, Harvard University

2:30 P.M. EVALUATION OF ECONOMIC REGULATION OF INDUSTRY[1] (Joint Session with Transportation and Public Utilities Group)
Chairman: GEORGE W. WILSON, Indiana University
Papers: FRED WESTFIELD, Vanderbilt University
Methodology of Evaluating Economic Regulation
BOYD NELSON, Federal Communications Commission
Costs and Benefits of Regulating Common Carrier Communications
ANN FRIEDLAENDER, Boston College
Costs of Economic Regulation of Rail Transport
Discussants: ALFRED E. KAHN, Cornell University
LARRY DARBY, Temple University
JAMES R. NELSON, Amherst College

2:30 P.M. FINANCIAL CONSTRAINTS AND MINORITY ECONOMIC DEVELOPMENT[6] (Joint Session with The American Finance Association)
Chairman: ANDREW F. BRIMMER, Federal Reserve Board of Governors
Papers: FRED E. CASE, University of California, Los Angeles
Financing for the Under-Houseed: Report on Research in Six Major Cities
WILFRED J. GARVIN, Small Business Administration
The Small Business Capital Gap: The Special Case of Minority Enterprise
HENRY S. TERRELL, Federal Reserve Board of Governors
Relative Wealth Accumulation of White and Black Households: The Statistical Evidence
EDWARD D. IRONS, Howard University and the National Bankers Association and ANDREW F. BRIMMER, Federal Reserve Board of Governors
The Black-Owned Banks: Two Assessments
Discussants: MARCUS ALEXIS, Northwestern University and University of California, Berkeley
WILLIAM G. GRIGSBY, University of Pennsylvania
ROBERT LINDSAY, JR., New York University

8:30 P.M. RICHARD T. ELY LECTURE[1]
Chairman: JAMES TOBIN, Yale University
Speaker: HARRY G. JOHNSON, University of Chicago and London School of Economics
The Keynesian Revolution and the Monetarist Counter-Revolution

Tuesday, December 29, 1970

8:30 A.M. THE POLITICAL ECONOMY OF ENVIRONMENTAL QUALITY[1]
Chairman: ROBERT HAVEMAN, University of Wisconsin
Paper: ALLEN V. KNEESE, Resources for the Future, Inc.
Environmental Pollution: Economics and Policy
Discussants: KENNETH BOULDING, University of Colorado
JACK W. CARLSON, U.S. Office of Management and Budget
RICHARD JUDY, University of Toronto
MARC J. ROBERTS, Harvard University

8:30 A.M. THE ECONOMICS OF POLITICAL DECENTRALIZATION[1] (Joint Session with The Public Choice Society)
Chairman: JAMES M. BUCHANAN, Virginia Polytechnic Institute
Papers: DAVID BRADFORD and WALLACE OATES, Princeton University
Tax Competition Among Local Jurisdictions
JAMES HEINS, University of Illinois
State-Local Response to Fiscal Decentralization
KENNETH O. KORTANEK and OTTO A. DAVIS, Carnegie-Mellon University
Centralization and Decentralization: The Political Economy of Public School Systems
Discussants: CHARLES J. GOETZ, Virginia Polytechnic Institute
THOMAS BORCHERDING, University of Washington
WALTER E. WILLIAMS, University of California, Los Angeles
JOHN G. HEAD, Dalhousie University

8:30 A.M. GRADUATE STUDENT PAPERS[6] (Joint Session with Omicron Delta Epsilon)
Chairman: JOHN D. GUILFOIL, New York University
Papers: (To be announced)

10:30 A.M. THE EXPERIMENTAL STUDY OF GRADUATED WORK INCENTIVES[1] (Joint Session with The Econometric Society)
Chairman: ROBERT LEVINE, The RAND Corporation
Paper: JACOB MINCER, Columbia University
The Experimental Study of Graduated Work Incentives
Discussants: HAROLD WATTS, University of Wisconsin
ALBERT REES, Princeton University
JAMES MORGAN, Survey Research Center, University of Michigan
GUY ORCUTT, Yale University
ROBERT HALL, Massachusetts Institute of Technology

10:30 A.M. COMMERCIAL POLICY AND LESS-DEVELOPED COUNTRIES[1]
Chairman: LAWRENCE KRAUSE, The Brookings Institution
Papers: BELA BALASSA, Johns Hopkins University and International Bank for Reconstruction and Development
Policies of Protection in Developing Countries
CHRISTOPHER CLAGUE, University of Maryland
Tariff Preferences and Separable Utility
INGO WALTER, University of Missouri
Nontariff Barriers and the Export Performance of Developing Economies
Discussants: HARALD B. MALMGREN, Overseas Development Council
ROBERT BALDWIN, University of Wisconsin
STEPHEN GUISINGER, Southern Methodist University

10:30 A.M. PROGRESS IN SOCIO-ECONOMIC ACCOUNTING[3] (Joint Session with The American Statistical Association)
Chairman: OTIS DUDLEY DUNCAN, University of Michigan
Papers: MANCUR OLSON, University of Maryland
Recent Progress in Social Accounts
RICHARD RUGGLES and NANCY RUGGLES, Yale University
Micro-Data and Macro-Accounts
ROSANNE COLE, International Business Machine Corporation
Appraising The Accuracy of National Income and Product Estimates

12:30 P.M. ECONOMIC OUTLOOK LUNCHEON (Joint with The American Finance Association and The American Statistical Association)
Chairman: MARTIN GAINSBRUGH, National Industrial Conference Board
Speaker: PAUL VOLCKER, Under-Secretary of the Treasury

2:30 P.M. THE STATE OF ECONOMICS: THE BEHAVIORAL AND SOCIAL SCIENCES SURVEY[1]
Chairman: WASSILY LEONTIEF, Harvard University
Papers: CHARLES SCHULTZE, The Brookings Institution
The Reviewers Reviewed
JOHN GURLEY, Stanford University
The State of Political Economy
Discussants: HENRY RIECKEN, Social Science Research Council
ROBERT M. SOLOW, Massachusetts Institute of Technology
ROBERT HEILBRONER, New School for Social Research

2:30 P.M. EDUCATIONAL PRODUCTION RELATIONSHIPS[1]
Chairman: FINIS WELCH, National Bureau of Economic Research
Papers: HERBERT GINTIS, Harvard University
Educational Production Functions and the Characteristics of Worker Productivity
ERIC A. HANUSHEK, United States Air Force Academy
Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-data
JOHN C. HAUSE, University of Minnesota, and National Bureau of Economic Research
Aptitude, Schooling, and Achievement as Determinants of Lifetime Earnings
Discussants: ALAN THOMAS, University of Chicago
ROGER ALCALY, Columbia University
CHARLES WILSON, University of California, Los Angeles

2:30 P.M. INTRA- AND INTER-GENERATION TRANSFERS (Joint Session with the Association for the Study of The Grants Economy)
Chairman: MARTIN PFAFF, Wayne State University
Papers: BENJAMIN BRIDGES, Social Security Administration
Tax-Transfer Programs: Current Redistributive Effects Among Age and Economic Status Classes
FREDERIC L. PRYOR, Swarthmore College
A Simulation Model of Inter-Generation Transfer of Wealth
A. LIEFMANN KEIL, University of the Saarland
Some Effects of Electron Cycles on Intra- and Inter-Generation Transfers
MOLLIE ORSHANSKY, Social Security Administration
Inter-Generation and Regional Transfers Through Public Income Maintenance

4:30 P.M. INVITED ADDRESS[1]
Chairman: TIBOR SCITOVSKY, Stanford University
Speaker: RAÚL PREBISCH, United Nations Economic Commission for Latin America
Development, Structural Changes, and International Cooperation

8:00 P.M. PRESIDENTIAL ADDRESS[2]
Chairman: PAUL H. DOUGLAS
Speaker: WASSILY LEONTIEF, Harvard University
Quantitative Analysis—Theoretical Assumptions and Nonobserved Facts

9:15 P.M. BUSINESS MEETING

Wednesday, December 30, 1970

8:30 A.M. MICRO ASPECTS OF MACRO PERFORMANCE[1]
Chairman: ARTHUR M. OKUN, The Brookings Institution
Papers: EDWARD BUDD and DAVID SEIDERS, Pennsylvania State University
The Impact of Inflation on the Distribution of Income and Wealth
ARNOLD PACKER, U.S. Office of Management and Budget
The Two-Way Relationships Between Budget Programs and Economic Activity
Discussants: CHARLES METCALF, University of Wisconsin
JOHN PALMER, Stanford University

8:30 A.M. THE THEORY OF PLANNING[1] (Joint Session with Association for Comparative Economics, The Association for the Study of Soviet-Type Economies, and The Econometric Society)
Chairman: EVSEY D. DOMAR, Massachusetts Institute of Technology
Papers: MICHAEL MANOVE, University of Michigan
A Model of Soviet-Type Economic Planning
RICHARD D. PORTES, Princeton University
Decentralized Planning Procedures and Centrally Planned Economies
Discussants: J. MICHAEL MONTIAS, Yale University
(Others to be named)

8:30 A.M. ECONOMIC THOUGHT[4]
Chairman: CRAUFURD D. GOODWIN, Duke University
Papers: KENNETH BOULDING, University of Colorado
After Samuelson, Who Needs Adam Smith?
WILLIAM GRAMPP, University of Illinois, Chicago
Classical Economics and Its Moral Critics
VINCENT TARASCIO, University of North Carolina, Chapel Hill
Some Recent Developments in the History of Economic Thought

10:30 A.M. THE VIETNAMESE WAR: ECONOMIC PROSPECTS AFTER DEMOBILIZATION (Joint Session with The Econometric Society)
Chairman: DANIEL SUITS, University of California, Santa Cruz
Paper: BERNARD UDIS, University of Colorado
The Vietnamese War: Economic Prospects After Demobilization

Discussants: MURRAY WEIDENBAUM, U.S. Treasury and Washington University, St. Louis
SANDOR FUCHS, University of California, Berkeley
LAWRENCE LYNN, Stanford University

10:30 A.M. GRANT ELEMENTS IN SOVIET-TYPE ECONOMIES (Joint Session with the Association for the Study of The Grants Economy)
Chairman: ROBERT W. CAMPBELL, Indiana University
Papers: PAUL JONAS, U.S. Agency for International Development, New Delhi, India
An International Comparison of Unilateral Transfers from Soviet-Type Economies
JAMES C. CARTER, University of Oregon
The Real Cost of Soviet Economic Aid
JANOS HORVATH, Butler University
Grant-Elements in Intra-Block Aid Programs
ALAN A. BROWN, Indiana University
Grant-Elements in the Hungarian National Accounts

10:30 A.M. CONGLOMERATE AND VERTICAL RESPONSES TO MARKET IMPERFECTION[1]
Chairman: RICHARD CYERT, Carnegie-Mellon University
Papers: JOHN LINTNER, Harvard University
Conglomerates and Finance: Recent Theory and Experience
OLIVER WILLIAMSON, University of Pennsylvania
The Vertical Integration of Production: A Market Failures Analysis
Discussants: ROLAND MCKEAN, University of Virginia
J. FRED WESTON, University of California, Los Angeles

10:30 A.M. NEW DIRECTIONS IN NATIONAL BUREAU RESEARCH—I[7] (One of two sessions, joint with The American Statistical Association, organized by The National Bureau of Economic Research to mark its fiftieth year, with papers by members of its staff)
Papers: WILLIAM LANDES
Law and Economics
FINIS WELCH
Human Resources Problems
JOHN R. MEYER and JOHN F. KAIN
Urban Economics Problems

2:30 P.M. TEACHING ECONOMICS: EXPERIMENTS AND RESULTS[1]
Chairman: RENDIGS FELS, Vanderbilt University
Papers: R. GRANN LLOYD, Tennessee State Teachers College
Teaching Economics to Black Students
PHILLIP SAUNDERS, Indiana University
The Lasting Effects of Elementary Economics Courses: Some Preliminary Results
BARRY CASTRO, Hostos Community College, New York
Experiments with Video Tape in Teaching Economics
Discussants: LESTER CHANDLER, Atlanta University
IRVING MORRISETT, University of Colorado
CAMPBELL MCCONNELL, University of Nebraska

2:30 P.M. EARNINGS DIFFERENTIALS[3] (Joint Session with The American Statistical Association)
Chairman: MELVIN W. REDER, Stanford University
Papers: VICTOR R. FUCHS, National Bureau of Economic Research
Male-Female Differentials in Hourly Earnings
LOWELL E. GALLAWAY, RICHARD K. VEDDER, and GENE L. CHAPIN, Ohio University
The Impact of Geographic Mobility on Regional Wage Differentials: A Test of the Steady State Equilibrium Hypothesis
WILLIAM R. BAILEY and ALBERT E. SCHWENK, Bureau of Labor Statistics
A Study of Wage Differentials Using Macro-Data from Wage Surveys
Discussants: THOMAS W. GAVETT, Bureau of Labor Statistics
MAHMOOD A. ZAIDI, University of Minnesota
GERALD W. SCULLY, Southern Illinois University

2:30 P.M. NEW DIRECTIONS IN NATIONAL BUREAU RESEARCH—II[7] (One of two sessions, joint with The American Statistical Association, organized by The National Bureau of Economic Research to mark its fiftieth year, with papers by members of its staff)
Papers: WARREN SANDERSON and ROBERT WILLIS
Population Growth and Distribution
THOMAS JUSTER
The Measurement of Social and Economic Performance
VICTOR ZARNOWITZ
Short-Term Forecasting

Announcements

Harvard University has a limited number of postdoctoral research fellowships available for the academic year 1971–72 for scholars engaged in research on aspects of Russian or East European life, particularily economics, government, history, literature, and sociology.

Scholars who already have financial support from another source may apply for the status of associate which carries no stipend but provides access to the research facilities of the Center and University at large. Applications are due by January 15, 1971. Please address inquiries to: Russian Research Center, 1737 Cambridge Street, Cambridge, Massachusetts 02138.

Journal of Development Planning

New publication on problems of development economics contains material prepared by the United Nations Secretariat and its consultants.

The *Journal* is published as part of the work program of the Centre for Development Planning, Projections and Policies of the department of economics and social affairs of the United Nations Secretariat. Inquiries may be addressed to Sales Section, United Nations, New York, N.Y. or Geneva, Switzerland.

The Asia Foundation has provided the American Economic Association with a fund to be used to assist students and visiting scholars from Asia to attend the annual meeting of the American Economic Association. The maximum amount available to an individual applicant is $100. Inquiries should be addressed to the AEA, 1313 21st Avenue, South, Nashville, Tennessee 37212.

A fourth *Summer Institute in Behavioral Science and Law* will be held in 1971 at the University of Wisconsin in Madison. The Institute is designed to train students in the methods and techniques of interdisciplinary legal research. Graduate students in the social science disciplines and law students are eligible to compete for stipends, travel allowances, and tuition waivers. For further information and applications, write to Professor Joel B. Grossman, Director, Center for Law and Behavioral Science, Social Science Building, University of Wisconsin, Madison, Wisconsin 53706.

The Association for the Study of the Grants Economy will hold its second annual symposium, to be chaired by its president Kenneth E. Boulding, jointly with the annual meetings of the American Association for the Advancement of Science (Chicago December 26–27, 1970), and with the American Economic Association and the American Statistical Association (Detroit, December 28–30, 1970).

Sessions will be held in Detroit on (1) The Theory and Effects of Implicit Grants, (2) Distributional Consequences of Public Choice Systems, (3) Transfers in Economic Development; and in Chicago on (4) The Grants Economy: Measures and Models, (5) Transfers as an Instrument of Ecological Policy, (6) Intra- and Intergeneration Transfers, (7) The Grant Economies in Comparative Perspective.

Economists and other social scientists who would like to contribute papers dealing with theoretical or empirical aspects of grants economies are invited to forward titles and abstracts of 300 words to the program coordinators: for the meetings of the AEA, Professor Martin Pfaff, 210 Prentis Bldg., Wayne State University, Detroit, Michigan 48202; for the meetings of AAAS and ASA, Professor Janos Horvath, Department of Economics, Butler University, Indianapolis, Indiana 46208.

The American Economic Association has received a further grant from the National Science Foundation to continue its *Visiting Scientist Program in Economics* during the 1970–71 academic year. The purpose of this program is to stimulate improved teaching and interest in modern economics at colleges and universities whose major focus is at the undergraduate level. Visits to about 40 campuses were provided by the program in 1969–70.

The NSF grant provides financing for visits to campuses by distinguished economists who spend a day or so on each visit. The program is planned jointly between the visitor and host institution; customarily it includes a talk to undergraduate major students, perhaps an informal seminar with faculty and students, discussion of current course and research interests of the host faculty, and other activities as may be desirable. While the ultimate concern is to encourage greater interest in, and better teaching of, economics at the undergraduate level, visitor's talks may be on research, current developments in economics, policy issues, or other topics that may seem appropriate to the needs of the particular institution.

This program is under the general direction of the Association's Committee on Economic Education. Professor Phillip Saunders administers the program under the policies established by the Committee. He works with an informal roster of economists around the country who may be available for such visits, depending on their own schedules and the degree of common interest between them and the inviting institutions. Institutions are free to suggest visitors they would especially like to have, so long as these are within a radius of 100–200 miles of the campus, in order to minimize travel costs. The host institution is expected to provide lodging and meals for the visitor when he is on campus; other costs will be covered by the grant.

Schools interested in applications should write to: Professor Phillip Saunders, Department of Economics, Indiana University, Bloomington, Indiana 47401.

The Financial Management Association

A group of twenty-five professors of business finance from universities in the United States and Canada met April 18, 1970 and voted to establish a professional association in the field of financial management. A

general organizational meeting will be held at the Christmas meetings of the Allied Social Science Association in Cobo Hall, Detroit at 4:30 P.M. on the first day of those meetings. All interested persons are invited to attend. Other sessions will be conducted at the meetings of the Western Finance Association in Davis, California on August 27–28, 1970; of the Southern Finance Association in Atlanta on November 13–14, 1970; and of the Appalachian Finance Asssociation in Montreal on April 16–17, 1971.

The purpose of the Financial Management Association is to serve the professional interests of those engaged in teaching, research, and corporate financial management. Preliminary plans call for annual meetings in the fall and publication of a journal entitled *Financial Management*. For more information, contact one of the members of the organization committee: Professor L. P. Anderson, College of Business, Colorado State University, Fort Collins, Colorado 80521; Professor H. Blythe, College of Business Administration, Ohio University, Athens, Ohio 45701; Professor E. F. Brigham, School of Business, University of Wisconsin, Madison, Wisconsin 53706; Professor J. W. Dunlap, College of Business Administration, University of Akron, Akron, Ohio 44304 (chairman); Professor M. R. Sussman, College of Business Administration, University of Georgia, Athens, Georgia 30602; Professor A. G. Sweetser, School of Business, State University of New York at Albany, Albany, New York 12203 (secretary).

Deaths

Norman Barcus, director, research and statistics division, Michigan Employment Security Commission, Jan. 13, 1970.

Carl Campbell, Portland State College, Feb. 15, 1970.

John H. Duckstad, Jan. 17, 1970.

Edmond S. Harris, Feb. 24, 1970.

Wallace F. Lovejoy, professor of economics, Southern Methodist University, Apr. 1, 1970.

John E. MacLeish, Aug. 3, 1969.

Charles F. Martell, Jr., Aug. 1969.

H. Austin Peck, State University of New York, Potsdam, Feb. 10, 1970.

(Mrs.) William Plunkert, June 12, 1969.

Jared Sparks, Jr., University of Arkansas, Feb. 6, 1970.

Lionel W. Thatcher, professor, University of Wisconsin, Madison, Apr. 26, 1970.

David Weintraub, Sept. 1969.

Harvey J. Wheeler, head, department of economics, Clemson University, May 14, 1970.

Louis B. Zapoleon, Dec. 27, 1969.

Retirements

G. Heberton Evans, Jr., Johns Hopkins University, June 1970.

David M. Harrison, professor of economics, Ohio State University, June 1970.

Alfred Kahler, professor of economics, New School for Social Research, June 1970.

Clement S. Logsdon, professor of marketing, University of North Carolina at Chapel Hill, May 31, 1970.

Glenn W. Miller, professor of economics, Ohio State University, Oct. 1970.

Walter F. Muhlbach, professor emeritus of finance, School of Business Administration, The American University, June 1970.

Allen V. Wiley, associate professor of economics, Bowling Green State University, Dec., 1969.

Clifford Zuroske, Washington State University, June 1970.

Visiting Foreign Scholars

Chris G. Archibald, University of Essex: department of economics, University of British Columbia, 1970–71.

Robert W. Daniels, University of Lancaster, England: professor of economics, Case Western Reserve University, July 1970.

William Gorman: visiting professor, department of economics, University of North Carolina, Chapel Hill.

Rene Lafon, Ecole Superior des Sciences Economiques et Commerciales, Paris: professor of international marketing programs, international business, George Washington University, 1970.

Raul Prebisch, Latin American Institute for Economic and Social Planning, Santiago, Chile: professor of Latin American studies, Columbia University, spring 1971.

Wilfred Prest, Melbourne University, Australia: visiting professor of economics, University of Pittsburgh, fall 1970.

A. V. R. Rao, Ministry of Development and Finance, Zambia, Lusaka: visiting professor of economics, Purdue University, spring and summer 1970.

E. B. A. St. Cyr, University of W. Indies: visiting professor of economics, Southern Illinois University, spring, 1970.

Promotions

F. Gerard Adams: professor, department of economics, University of Pennsylvania.

James G. Allgood: extension associate professor of economics, North Carolina State University.

Glen W. Atkinson: associate professor of economics, University of Nevada, July 1970.

Richard A. Barrett: associate professor of business administration, George Washington University.

C. Phillip Baumel: professor of economics, Iowa State University, July 1970.

David E. Bond: associate professor, department of economics, University of British Columbia.

Karl Bonutti: associate professor, department of economics, Cleveland State University, Sept. 1970.

J. Hayden Boyd: associate professor of economics, Ohio State University.

William Breit: professor, department of economics, University of Virginia, Sept. 1970.

John Burke: associate professor, department of economics, Cleveland State University, Sept. 1970.

Barry R. Chiswick: associate professor, Columbia University.

Shih-Fan Chu: associate professor of economics, University of Nevada, July 1970.

Norris C. Clement: associate professor, department of economics, San Diego State College.

George J. Conneman, Jr.: professor of agricultural economics, Cornell University, July 1970.

Stanislaw Czamanski: professor, department of city and regional planning, Cornell University.

J. Ronnie Davis: associate professor of economics, Iowa State University, Sept. 1970.

Peter A. Diamond: professor of economics, Massachusetts Institute of Technology.

Loraine Donaldson: professor of economics, Georgia State University.

Dennis J. Dugan: associate professor of economics, University of Notre Dame, fall 1970.

Robert M. Dunn, Jr.: associate professor, department of economics, George Washington University, Sept. 1970.

Michael R. Edgmand: associate professor of economics, Oklahoma State University.

Edwin J. Elton: associate professor of finance, New York University.

Gerald I. Eyrich: assistant professor of economics, Claremont Men's College.

Marvin Fischbaum: associate professor of economics, Indiana State University.

Gene A. Futrell: professor of economics, Iowa State University, July 1970.

Donald Garnel: professor of economics, San Jose State College.

Dana C. Goodrich: professor of agricultural economics, Cornell University, July 1970.

Michael J. Greenwood: associate professor, department of economics, Kansas State University, July 1970.

K. L. Gupta: associate professor of economics, University of Alberta, 1970.

Josef Hadar: professor of economics, Case Western Reserve University.

Geoffrey B. Hainsworth: associate professor, department of economics, University of British Columbia.

DeVerle P. Harris: associate professor of mineral economics, Pennsylvania State University, July 1969.

John R. Harris: associate professor of economics, Massachusetts Institute of Technology.

Clark A. Hawkins: professor of economics and finance, University of Arizona.

Edward S. Howle: associate professor of economics, University of North Carolina at Chapel Hill.

Donald L. Huddle: professor of economics, Rice University.

Loren A. Ihnen: professor of economics, North Carolina State University.

Hirschel Kasper: professor of economics, Oberlin College.

Peter Kilby: associate professor, Wesleyan University.

Arthur B. Laffer: associate professor of business economics, Graduate School of Business, University of Chicago.

Gerald M. Lage: associate professor of economics, Oklahoma State University.

Wilford L. L'Esperance: professor of economics, Ohio State University.

Charles H. Little: associate professor of experimental statistics and economics, North Carolina State University.

Bennett T. McCallum: associate professor, department of economics, University of Virginia, Sept. 1970.

John McEntaffer: associate professor of economics, Indiana State University.

G. S. Maddala: professor of economics, University of Rochester, May 1970.

Murugappa C. Madhavan: associate professor, department of economics, San Diego State College.

Richard G. Marcis: associate professor of economics, Bowling Green State University.

Stanley H. Masters: associate professor of economics, department of economics, Rutgers–The State University.

Leo V. Mayer: associate professor of economics, Iowa State University, July 1970.

Jagdish C. Mehra: associate professor, department of economics, Youngstown State University.

William C. Merrill: professor of economics, Iowa State University, Sept. 1970.

Walther P. Michael: associate professor of economics, Ohio State University.

Nicholas A. Michas: associate professor, department of economics, Loyola University of Chicago.

John H. Moore: associate professor, department of economics, University of Virginia, Sept. 1970.

Hosseine Morewedge: associate professor of economics, Long Island University, Feb. 1970.

Rodney J. Morrison: associate professor of economics, Wellesley College.

Keizo Nagatani: associate professor, department of economics, University of British Columbia.

William G. Nelson: director, purchases and central services division, University of Pittsburgh, May 1970.

Richard T. Newcomb: associate professor of mineral economics, Pennsylvania State University, July 1969.

T. Everett Nichols, Jr.: extension professor of economics, North Carolina State University.

William H. Oakland: associate professor of political economy, Johns Hopkins University.

Hans C. Palmer: associate professor of economics, Claremont Graduate School, Pomona College.

Alfred Parker: associate professor of economics, University of New Mexico.

Rudolph G. Penner: professor of economics, University of Rochester, May 1970.

Michael J. Piore: associate professor of economics, Massachusetts Institute of Technology.

Martin F. J. Prachowny: associate professor of economics, Queen's University.

Sherwin Rosen: professor of economics, University of Rochester, May 1970.

Murray N. Rothbard: professor of economics, Polytechnic Institute of Brooklyn.

Edward V. Roy: associate professor, department of economics, State University of New York, Stony Brook.

H. Arthur Sandman: assistant professor of economics, North Carolina State University.

Robert L. Sandmeyer: professor of economics, Oklahoma State University.

Ronald A. Shearer: professor of economics, University of British Columbia.

Karl Shell: professor, department of economics, University of Pennsylvania.

J. Marvin Skadberg: associate professor of economics, Iowa State University, July 1970.

Gary W. Sorenson: associate professor, department of economics, Oregon State University.

Frank G. Steindl: professor of economics, Oklahoma State University.

James A. Stephenson: professor of economics, Iowa State University, Sept. 1970.

Paul Swadener: associate professor of business administration, College of Business Administration, University of Oregon, July 1970.

Thomas R. Swartz: associate professor of economics, University of Notre Dame, fall 1970.

Richard E. Sylla: associate professor of economics, North Carolina State University.

Benjamin J. Taylor: professor of economics, Arizona State University.

Peter Temin: professor of economics, Massachusetts Institute of Technology.

Lloyd B. Thomas, Jr.: assistant professor, department of economics, Kansas State University, July 1970.

Lester C. Thurow: professor of economics, Massachusetts Institute of Technology.

William C. Tomek: professor of agricultural economics, Cornell University, July 1970.

Hiroki Tsurumi: associate professor of economics, Queen's University.

Edwin F. Ulveling: associate professor of economics, Georgia State University.

R. Charles Vars, Jr.: associate professor, department of economics, Oregon State University.

Charles Waldauer: associate professor of economics, Pennsylvania Military Colleges.

Roman L. Weil, Jr.: associate professor of mathematical economics, Graduate School of Business, University of Chicago.

James Willis: associate professor of economics, San Jose State College.

Charles Wishart: assistant professor of economics, Indiana State University.

Murray Wolfson: professor, department of economics, Oregon State University.

Alexej Wynnyczuk: professor, department of economics, Rensselaer Polytechnic Institute.

Administrative Appointments

Mustapha K. Baksh: chairman, department of economics, Belknap College.

John O. Blackburn: provost, Duke University.

Howard R. Bowen: president, Claremont Graduate School and University Center.

Shun-hsin Chou: acting chairman, department of economics, University of Pittsburgh, Jan-Aug. 1970.

Frank A. Close: chairman, department of economics, Clemson University, July 1970.

David G. Davies: chairman, department of economics, Duke University.

Robert Evans, Jr.: chairman, department of economics, Brandeis University.

Daniel O. Fletcher: acting chairman, department of economics, Denison University, 1970–71.

Irving J. Goffman: chairman, department of economics, University of Florida, July 1970.

Elmer R. Gooding: assistant dean, College of Business Administration, Arizona State University.

Carl W. Hale: chairman and associate professor of economics, Auburn University, July 1970.

William L. Henderson: assistant to the president of the college, Denison University, June 1, 1970.

M. Bruce Johnson, chairman, department of economics, University of California at Santa Barbara, July 1970.

Peter B. Kenen: provost, Columbia University, Aug. 1969.

Kelvin J. Lancaster: chairman, department of economics, Columbia University, July 1970.

G. Peter Lauter: professor of business administration; director, Programs in International Business, The George Washington University.

Leonard Lecht, National Planning Association: chairman, department of economics, Cleveland State University, Sept. 1970.

Larry C. Ledebur: associate dean of men, Denison University, Aug. 1970.

Carl Liedholm: chairman, department of economics, Michigan State University, June 1970.

Matityahu Marcus: professor and chairman, department of economics; director, Bureau of Economic Research, Rutgers-The State University.

Marlen F. Miller, Washington State University: chairman, department of economics, Pacific Lutheran University.

Reuben G. Miller, Smith College: chairman, department of economics, Sweet Briar College, Sept. 1970.

Francis T. O'Brien: acting chairman, department of economics, Providence College, 1970–71.

Douglas S. Paauw, National Planning Association: chairman, department of economics, Wayne State University.

Harry R. Page: chairman, department of business administration, George Washington University, July 1970.

John A. Shaw: associate professor, chairman, department of economics, Fresno State College.

Donald R. Sherk: chairman, department of economics, Simmons College, July 1970.

J. Graham Smith: chairman, department of economics, McGill University.

Donald L. Sternitzke: chairman, department of economics, Bowling Green State University, Aug. 1970.

Richard S. Wallace, Georgia State University: dean, School of Business Administration, Winthrop College, July 1970.

Jack C. Wimberly: acting chairman, department of economics, University of Southern Mississippi.

New Appointments

Elsayed A. Abdou: instructor of business administration, University of Tennessee at Chattanooga.

Glen G. Alexandrin: assistant professor of economics, Villanova University, fall 1970.

Polly R. Allen, New York University: assistant professor, department of economics, Princeton University.

Frederick Amling, University of Rhode Island: professor of finance, George Washington University.

Robert E. Anderson, Johns Hopkins University: instructor, department of economics, University of Illinois.

Robert Ankli, Guelph University: visiting assistant professor, department of economics, University of Illinois.

Gary Appel, Michigan State University: assistant professor of economics, Georgia State University.

K. L. Avio: lecturer, department of economics, University of Western Ontario, July 1970.

Bennett Baack, University of Washington: assistant professor, department of economics, Ohio State University.

Nancy Baggott, Purdue University: assistant professor, Wayne State University.

Elba F. Baskin: lecturer in accounting, School of Business Administration, University of North Carolina at Chapel Hill.

H. G. Baumann: lecturer, department of economics, University of Western Ontario, July 1970.

William Baumol: visiting professor, department of economics, Graduate School, New York University.

Richard Bean: assistant professor, department of economics, University of Houston, spring 1970.

Steven Beggs: instructor, department of economics, Boston College.

Donald M. Bellante: assistant professor, department of economics and geography, Auburn University, Sept. 1970.

Neville Bembridge: instructor in economics, Lafayette College, Sept. 1970.

James T. Bennett: assistant professor of economics, George Washington University, Sept. 1970.

Ramesh Bhardwaj: associate professor, department of economics, Cleveland State University, Sept. 1970.

Roger D. Blair: assistant professor of economics, University of Florida, 1970–71.

Allen Blitstein: assistant professor, department of economics, University of Arizona.

Douglas Bohi, Caterpillar Tractor, Peoria: assistant professor of economics, Southern Illinois University.

Christine H. Branson: instructor in economics, Simmons College.

Michael J. Brennan, Massachusetts Institute of Technology: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

D. L. Brito: assistant professor, department of economics, University of Wisconsin, Madison.

Frederick J. Brooks-Hill, University of Pennsylvania: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

William R. Buechner, Harvard University: assistant professor, department of economics, Smith College.

George W. M. Bullion, Purdue University: economist, U.S. Department of Agriculture.

Robert W. Butler, Jr., Instituto Tecnologico Autonomo de Mexico: assistant professor of economics, Saint Louis University.

William L. Call, Northwestern State College: assistant professor of economics, Kanawha Valley Graduate Center, West Virginia University, Sept. 1970.

Joel Caron: lecturer, department of economics, University of Illinois.

James H. Carrington: associate professor of business management, School of Business Administration, American University, 1970–71.

Richard Cebula: instructor of economics, Georgia State University.

John K. Chang, Lafayette College: industrial economist, Asian Development Bank, Manila, Phillippines.

Albert Church: assistant professor of economics, University of New Mexico, fall 1970.

Robert W. Clarke: associate professor of accounting, School of Business Administration, University of North Carolina at Chapel Hill.

William A. Clarke: lecturer, department of economics, Rutgers–The State University.

Michael Claudon, Johns Hopkins University: instructor, department of economics, Middlebury College.

Thomas W. Cleaver: assistant professor of economics, Villanova University, fall 1970.

Warren Coats, University of Chicago: acting assistant professor of economics, University of Virginia, Sept. 1970.

William Conrad: assistant professor of economics, Georgia State University.

Frank Corcione, Lehigh University: instructor in economics, Lafayette College, Sept. 1970.

John Cornwall, Tufts University: professor, department of economics, Southern Illinois University.

Steven R. Cox: assistant professor of economics, Arizona State University.

Virginia Cross: visiting assistant professor of business administration, University of Tennessee at Chattanooga.

Christopher Curran: assistant professor of economics, School of Business Administration, Emory University, Sept. 1970.

Donald Curran: associate professor, department of economics, Cleveland State University, Sept. 1970.

Camilo Dagum, University of Mexico: visiting professor, department of economics, University of Iowa.

Raymond Daniel, Purdue University: assistant professor, department of agricultural economics, University of Tennessee, Mar. 1970.

David Dannenbring: lecturer, School of Business Administration, University of North Carolina at Chapel Hill.

Michael R. Darby, University of Chicago: assistant professor, department of economics, Ohio State University.

Partha Dasgupta: assistant professor, department of economics, Carnegie-Mellon University, May 1970.

George de Menil, Boston College: assistant professor, department of economics, Princeton University.

David A. Denslow: assistant professor of economics, University of Florida, 1970–71.

Paul R. Deuster: assistant professor of economics, Ohio University, Sept. 1970.

Albert S. Dexter, Columbia University: assistant professor, Commerce and Business Administration, University of British Columbia, fall 1970.

T. J. O. Dick; lecturer, department of economics, University of Western Ontario, July 1970.

Paul F. Dickens: assistant professor of economics, Arizona State University.

Erwin Diewert: associate professor, department of economics, University of British Columbia.

Larry Dildine: assistant professor of economics, Georgia State University.

Dale W. Dison: assistant professor, department of economics and geography, Auburn University, Sept. 1970.

Salvatore Divita: associate professor of business administration, George Washington University.

William Drazen: Instructor in economics, Skidmore College.

Daniel M. Driscoll: assistant professor of economics, Providence College, fall 1970.

H. C. Driver, University of Manitoba: associate professor, department of agricultural economics, University of Guelph.

John E. Drotning: visiting professor of Industrial Relations Research Institute, University of Wisconsin.

Lawrence A. Duewer, U.S. Department of Agriculture: assistant professor, agricultural economics, Purdue University.

Louis Ederington: assistant professor of economics, Georgia State University.

Ronald Ehrenberg: assistant professor, department of economics, Loyola University of Chicago, 1970–71.

Alan Eliason, University of Minnesota: assistant professor of management, College of Business Administration, University of Oregon, Sept. 1970.

Leon M. Ennis, Jr.: assistant professor of economics, North Carolina State University.

Dennis Erickson: assistant professor of economics, University of Tennessee at Chatanooga.

John R. Eriksson: economist, Office of Program and Policy Coordination, U.S. Agency for International Development.

Ibrahim Eris: assistant professor of economics, Rice University.

Ulrich F. Ernst, Indiana University: assistant professor, department of economics, Ohio State University.

John L. Evans, University of North Carolina: assistant professor, Commerce and Business Administration, University of British Columbia, fall 1970.

Donald Farness, Fresno State College: associate professor, department of economics, Oregon State University, fall 1970.

Richard A. Farrar: professor of economics, University of Evansville, 1970.

Ronald Findlay, University of Rangoon: professor of economics, Columbia University, 1970–71.

Arthur Ford, New School for Social Research: assistant professor of economics, Southern Illinois University.

John E. Fredland: assistant professor of economics, George Washington University, Sept. 1970.

Bernard Friedman: lecturer and assistant provost for planning, department of economics, Columbia University.

Winship C. Fuller: assistant professor of economics, Providence College, fall 1970.

Monique Garrity: assistant professor of economics, Wellesley College.

Robert J. Gayton, University of California, Berkeley: assistant professor, Commerce and Business Administration, University of British Columbia, Jan. 1971.

Peter T. Gottschalk: assistant professor of economics, Williams College.

Andrew W. Green, University of Maryland: associate professor of economics, West Chester State College, Sept. 1970.

Donald W. Green, University of California at Berkeley: assistant professor, department of economics, University of Pennsylvania, fall 1970.

Clyde R. Greer: assistant professor, department of agricultural economics, Montana State University, Feb. 1970.

W. Smith Greig: professor, department of agricultural economics, Washington State University.

Gene Gruver, Iowa State University: assistant professor, University of Pittsburgh.

Charles E. Guy: assistant professor, department of economics and geography, Auburn University, Sept. 1970.

Andrew Hamer, University of Massachusetts: assistant professor of economics, Georgia State University.

Ronald W. Hansen: assistant professor, department of economics, University of Notre Dame, fall 1970.

Thomas M. Havrilesky, Duke University: visiting associate professor of economics, Rice University.

Delbert I. Hawkins, Southern Illinois University: assistant professor of marketing, College of Business Administration, University of Oregon, Sept. 1970.

Pamela J. Heath: assistant professor of economics, University of Denver.

Robert F. Hebert, Louisiana State University: assistant professor of economics, Clemson University, Aug. 1970.

Donald G. Heckerman: associate professor of economics, University of Arizona.

James J. Heckman, Princeton University: lecturer, department of economics, Columbia University, 1970–71.

Michael A. Heilperin: visiting professor, Hillsdale College, 1970–71.

Mary A. Hines: assistant professor of business administration, Butler University, Sept. 1970.

Samuel Ho: associate professor, department of economics, University of British Columbia.

Timothy D. Hogan: assistant professor of economics, Arizona State University.

Marjorie H. Honig: lecturer in economics, Douglass College, Rutgers–The State University.

Dennis K. Hoover, instructor, chairman, extension economics, University of Wisconsin.

David L. Horner, University of Wisconsin: assistant professor, department of economics, Wayne State University.

Yashuhisa Hosomatsu: assistant professor of economics, Georgia State University.

Carl B. Housley: research associate and assistant professor of economics, University of Southern Mississippi.

Yukon Huang, Princeton University: acting assistant professor of economics, University of Virginia, Sept. 1970.

Michael Hudson: assistant professor of economics, New School for Social Research, fall 1970.

Robert Hume: lecturer, department of economics, University of Illinois.

Stephen Hymer: professor of economics, New School for Social Research, fall 1970.

Robert P. Inman, Harvard University: assistant professor, department of economics, University of Pennsylvania, fall 1970.

Henry E. Jakubiak: assistant professor of economics, Rutgers–The State University.

Kenneth Jameson: assistant professor, department of economics, University of Notre Dame, fall 1970–71.

Christopher Jehn: assistant professor of economics, George Washington University, Sept. 1970.

L. Todd Johnson: assistant professor of accounting, Rice University.

William G. Johnson: research associate, bureau of economic research, Rutgers–The State University.

Charles P. Jones, Indiana University: assistant professor of economics, North Carolina State University.

James L. Joy, University of Denver: assistant professor, department of economics, University of Nebraska at Omaha.

Carol L. Jusenius, New York University: assistant professor, department of economics, Smith College.

Milton Z. Kafoglis: professor of economics, University of Florida, 1970–71.

George Kaufman, University of Southern California: professor of banking, College of Business Administration, University of Oregon, Sept. 1970.

Harry H. Kelejian: visiting associate professor, department of economics, Graduate School, New York University.

Harvey Kerachsky: lecturer in finance, department of economics, New York University.

Joseph A. Kershaw: professor of economics, Williams College.

Basheer A. Khumawala: lecturer, School of Business Administration, University of North Carolina at Chapel Hill.

Young C. Kim: assistant professor of economics, Northern Illinois University.

Robert S. Kline: associate professor, School of Business Administration, Winthrop College.

Tetsunori Koizumi, Brown University: assistant professor, department of economics, Ohio State University.

Ira L. Koslow: visiting lecturer, department of economics, California State College, Long Beach, 1970–71.

John Kraft: assistant professor of economics, University of Florida, 1970–71.

David Kresge, Harvard University: associate professor of economics and finance, New York University.

Edward B. Krinsky: visiting assistant professor of industrial relations and economics, University of Wisconsin.

Kazuhisa Kudoh, University of Chicago: assistant professor, department of economics, San Diego State College.

Robert G. Kuller, University of Kansas: assistant professor of economics, University of Wyoming.

Conway L. Lackman: instructor, department of economics, School of Business Administration, The American University, 1970–71.

John C. Lambelet, Harvard University: assistant professor, department of economics, University of Pennsylvania, fall 1970.

John S. Lapp: assistant professor of economics, North Carolina State University.

Joseph F. Lee: associate professor, School of Business Administration, University of North Carolina at Chapel Hill.

Kyu S. Lee, New York University: assistant professor, department of economics, Wayne State University.

Thomas M. Lenard, Brown University: assistant professor, department of economics, University of California, Davis.

Eddie M. Lewis: instructor, department of economics, University of Southern Mississippi.

Joseph A. Licari: instructor in economics, Occidental College.

Russel Lidman: instructor in economics, Oberlin College.

Peter C. Lin: assistant professor of economics and business, University of Alaska.

Cliff Lloyd, University of New York at Buffalo: professor of economics, University of Iowa.

Thomas Lyon: assistant professor of economics, University of Tennessee at Chattanooga.

Kenneth R. MacCrimmon, Carnegie-Mellon University: professor, commerce and business administration, University of British Columbia, fall 1970.

Michael B. McElroy: assistant professor of economics, North Carolina State University.

Paul McGouldrick, McGill University: associate professor of economics, State University of New York, Binghamton.

Robert McGuckin: acting assistant professor, department of economics, University of California, Santa Barbara.

Fritz Machlup: visiting professor, department of economics, Graduate School, New York University.

Robert K. Main: associate professor, department of economics and geography, Auburn University, June 1970.

J. Robert Malko: assistant professor of economics, business and economics division, Illinois Wesleyan University.

Don H. Mann, California State College: instructor,

department of economics and commerce, Simon Fraser University.

Judith K. Mann, San Fernando Valley State College: instructor, department of economics and commerce, Simon Fraser University.

Donald J. Mathieson, Stanford University: lecturer, department of economics, Columbia University, 1970–71.

Wolfgang Mayer: assistant professor of economics, University of Cincinnati, Sept. 1970.

Thomas H. Mayor: associate professor, department of economics, University of Houston, summer 1969.

Allan I. Meindelowitz: lecturer, department of economics, Rutgers-The State University.

Jerome W. Milliman, Indiana University: director, Center for Urban Affairs; professor of economics, University of Southern California.

Edwin S. Mills, Johns Hopkins University: professor of economics and public affairs, Princeton University.

Oskar Morgenstern, Princeton University: professor, department of economics, Graduate School, New York University.

Brian Motley: visiting associate professor of economics, University of Rochester, Sept. 1970.

Tridib K. Mukherjee, Fairmont State College: assistant professor of economics and business, Appalachian State University.

Badal Mukhopadhyay: assistant professor of political economy, Johns Hopkins University, 1970–71.

Robert H. Nelson, The City College: assistant professor, department of economics, Uptown University, Sept. 1970.

David M. Nienhaus: assistant professor, department of economics, State University of New York, Stony Brook.

Normand R. Noel: assistant professor of economics, Providence College.

Louis Noyd: instructor of economics, University of Tennessee at Chattanooga.

Wallace Oates: visiting associate professor, department of economics, Graduate School, New York University.

Edward O'Connor: instructor of business administration, University of Tennessee at Chattanooga.

James F. O'Connor: assistant professor, department of economics, Virginia Polytechnic Institute.

Jonathan Ogur, Cornell University: assistant professor, department of economics, Tulane University, Sept. 1970.

Donald J. O'Hara, University of California, Los Angeles: assistant professor of economics, University of Rochester, Sept. 1970.

Edgar Olsen, The RAND Corporation: assistant professor of economics, University of Virginia, Sept. 1970.

Douglas Olson, University of California, Los Angeles: assistant professor of quantitative methods, College of Business Administration, University of Oregon, Sept. 1970.

Kent W. Olson: visiting assistant professor of economics, Oklahoma State University.

A. Desmond O'Rourke, University of California, Davis: assistant professor, department of agricultural economics, Washington State University.

Kristian S. Palda: visiting professor, School of Business, Queen's University, July 1970.

Jacob Paroush: visiting associate professor, department of economics, Rutgers-The State University.

Donald Parsons, University of Chicago: assistant professor, department of economics, Ohio State University.

Peter Passell, Yale University: lecturer, department of economics, Columbia University, 1970–71.

Svetozar Pejovich, Texas A.&M. University: professor of economics, Ohio University, Sept. 1970.

Nicholas S. Perna: economist, research department, Federal Reserve Bank of New York.

Bruce L. Petersen, National Bureau of Economic Research: assistant professor of economics, University of Wyoming.

Mohan V. Phatak, University of Rochester: assistant professor, department of economics, San Diego State College.

Richard W. Pollay, University of Kansas: associate professor, commerce and business administration, University of British Columbia, fall 1970.

Clayne L. Pope, University of Michigan: assistant professor of economics, Brigham Young University.

Wolfhard Ramm, Northwestern University: assistant professor, department of economics, University of California, San Diego.

Edward J. Ray, Stanford University: assistant professor, department of economics, Ohio State University.

Alex Rebmann-Huber: visiting assistant professor, department of economics, University of British Columbia.

Michael Redisch, Northwestern University: assistant professor of economics, University of Iowa.

Eleanor Reibstein: instructor, department of economics, Adelphi University.

David J. Reid: lecturer, department of economics and commerce, Simon Fraser University.

Jack Rich: assistant professor, department of economics, Oregon State University, fall 1970.

James A. Richardson, University of Michigan: assistant professor, economics department, Louisiana State University, Sept. 1970.

J. David Richardson: assistant professor, department of economics, University of Wisconsin, Madison.

Blaine Roberts: assistant professor economics, University of Florida, 1970–71.

James G. Robertson: assistant professor, department of economics, Bowling Green State University, Sept. 1970.

Donald K. Rose: assistant professor, department of agricultural economics, Montana State University, Sept. 1970.

Hugh Rose, Harvard University: professor of political economy, Johns Hopkins University.

Steven S. Rosefielde, lecturer, department of economics, University of North Carolina at Chapel Hill.

Daniel Rubinfeld: instructor of economics, Wellesley College.

Ronald J. Rudolf: lecturer, department of economics, Rutgers–The State University.

Lawrence A. Rupley, Knox College: lecturer in economics, Shmadu Bello University, Zaria, Nigeria.

Frederick A. Russ: lecturer, School of Business Administration, University of North Carolina at Chapel Hill.

Keith P. Russell: assistant professor, department of economics, Virginia Polytechnic Institute.

David J. S. Rutledge: acting assistant professor, Food Research Institute, Stanford University, 1970–71.

Ajit Sabharwal, The City College: instructor, department of economics, Uptown University, Sept. 1970.

Lars Sandberg, Harvard University: associate professor, department of economics, Ohio State University.

Rubin Saposnik, University of Kansas: professor of economics, Georgia State University.

Bernard Sarachek: visiting professor of economics, Rutgers-The State University.

Barbara A. Sawtelle: assistant professor of economics, Simmons College.

Richard Scheffler: visiting lecturer, department of economics, University of North Carolina at Chapel Hill.

David Scheffman: instructor, department of economics, Boston College.

Richard Schmalensee, Massachusetts Institute of Technology: assistant professor, department of economics, University of California, San Diego.

Larry Schroeder: assistant professor of economics, Georgia State University.

David L. Schulze: assistant professor of economics, University of Florida, 1970–71.

James W. Seal: instructor, department of economics, University of Southern Mississippi.

Ahmad H. Shamseddine, Elizabethtown College: associate professor of economics, West Chester State College, Sept. 1970.

Edward Shapiro, University of Toledo: visiting lecturer, department of economics, California State College, Long Beach, 1970–71.

James C. Sharf: instructor, industrial relations and personnel management, School of Business Administration, The American University, 1970–71.

William F. Sharpe, University of California, Irvine: professor of business, Graduate School of Business, Stanford University.

Robert B. Shelton: assistant professor of economics, Arizona State University.

David Sjoquist, College of Saint Thomas: assistant professor of economics, Georgia State University.

Peter E. Sloane, Westfield State College: associate professor in economics and education, Clark University.

Dean C. Smith: visiting associate professor of economics, University of Arizona.

Richard K. Smith, Muskingum College: instructor, department of economics and business administration, Marietta College, Sept. 1970.

William G. Snead: assistant professor of economics, Hamilton College, July 1970.

Richard J. Solie: professor of economics and business, University of Alaska.

Lewis C. Solmon, assistant professor, department of economics, City University of New York, Sept. 1970.

Clive Southey: visiting assistant professor, department of economics, University of British Columbia.

Mitchell Stengel, Harvard University: instructor of economics, Mighigan State University, Sept. 1970.

Marcia Stigum: assistant professor, department of economics, Loyola University of Chicago, 1970–71.

P. A. V. B. Swamy, University of Wisconsin: associate professor, department of economics, Ohio State University.

Gerald J. Swanson: assistant professor, department of economics, University of Arizona.

Cary Swoveland, University of California, Los Angeles: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

Leslie Szeplaki, University of California, Riverside: assistant professor of economics, Northern Illinois University.

Michael Tannen, National Bureau of Economic Research: lecturer, department of economics, State University of New York at Binghamton.

David Tarr, Brown University: assistant professor, department of economics, Ohio State University.

Nancy Teeters: research associate, The Brookings Institution, Economic Studies Program, Feb. 1970.

Irwin Tepper, University of Pennsylvania: assistant professor, commerce and business administration, University of British Columbia, Jan. 1971.

Angelos A. Tsaklanganos: lecturer, department of economics, Rutgers-The State University.

Takashi Tsushima, Johns Hopkins University: assistant professor of economics, University of Alberta, 1970–71.

Paul J. Uselding, University of Illinois: assistant professor of political economy, Johns Hopkins University, 1970–71.

Dean H. Uyeno, Northwestern University: assistant professor, commerce and business administration, University of British Columbia, Jan. 1971.

Charles D. Van Eaton, Northwestern State College: associate professor of economics, Western Kentucky University, Sept. 1970.

M. Veeman, University of California at Berkeley: assistant professor, agricultural economics, University of Alberta, 1970–71.

T. Veeman, University of California at Berkeley: assistant professor of economics, University of Alberta, 1970–71.

Sundararajan Venkataraman, Harvard University: instructor in economics, University College of New York University, Sept. 1970.

Philip Verleger: acting assistant professor, department of economics, University of California, Santa Barbara.

Ilan Vertinsky, Northwestern University: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

Gottfried Voelker: assistant professor of economics, University of Tennessee at Chattanooga.

Richard F. Wacht: visiting associate professor of

business administration, University of North Carolina at Chapel Hill.

Paul A. Wachtel, instructor, department of economics, City University of New York, Sept. 1970.

Gordon A. Walter, University of California, Berkeley: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

Ingo Walter, University of Missouri-Saint Louis: professor of economics and finance, New York University.

Stanley E. Warner: assistant professor of economics, University of Florida, 1970–71.

Fumihiko Watanabe, University of Utah: assistant professor, department of economics, University of Nebraska at Omaha.

Joseph Waters, Cornell University: instructor, economics department, Middlebury College.

William G. Waters: assistant professor, department of economics, University of British Columbia, fall 1970.

John K. Welsby: assistant professor, commerce and business administration, University of British Columbia, fall 1970.

John T. Wenders: visiting associate professor, department of economics, University of Arizona.

Lawrence J. White, Ford Foundation, Indonesia: assistant professor, department of economics, Princeton University.

C. Robert Wichers: assistant professor, department of economics, State University of New York, Stony Brook.

Jim Wiebe: assistant professor of economics, South Dakota State University, May 1970.

John Wile, Brown University: instructor, department of economics, Rensselaer Polytechnic Institute, fall 1970.

Benjamin Wolkowitz, Brown University: assistant professor, department of economics, Tulane University, Sept. 1970.

Frank Zahn, Jr., University of Houston: assistant professor, department of economics, Bowling Green State University, Sept. 1970.

Paul M. Zipin, University of Connecticut: assistant professor of economics, Clemson University, Aug. 1970.

Barbara S. Zoloth: assistant professor of economics, School of Business Administration, Emory University, Jan. 1971.

Leaves for Special Appointments

Peter S. Barth, Ohio State University: Economic olicy Fellowship, The Brookings Institutions.

J. Hayden Boyd, Ohio State University: associate ofessor, University of Rochester.

Julian Buckley, New York University: controller, City of New York, 1970–71.

John G. Cragg, University of British Columbia: research director, Prices and Incomes Commission, Otwa.

awrence B. Darrah, Cornell University: visiting fessor, department of agricultural economics, Unisity of the Philippines, 1970–72.

aren Davis, Rice University: Research Fellowship, Brookings Institution.

Louis A. Dow, University of Houston: visiting professor, department of economics, University of Alabama, 1970–71.

Richard A. Easterlin, University of Pennsylvania: professor, Center for Behavioral Sciences, Stanford University, 1970–71.

Kenneth G. Elzinga, University of Virginia: economic assistant to the Assistant Attorney General, Anti-Trust Division, Department of Justice.

Olan D. Forker, Cornell University: USAID Mission, Turkey.

Charles E. Gearing, University of North Carolina at Chapel Hill: visiting professor, Middle East Technical University, Ankara, Turkey.

Richard C. Gerhan, Baldwin-Wallace College: System Development Corporation, Santa Monica, California.

Thomas H. Hibbard, Middlebury College: visiting assistant professor, department of economics and commerce, Simon Fraser University.

Dale M. Hoover, North Carolina State University: visiting professor of agricultural economics, University of California at Davis.

Harris E. Hordon, Jersey City State College: Economic Policy Fellow, The Brookings Institution, 1970–71.

Martin O. Ijere, Claremont Men's College: associate professor of economics, University of Nigeria.

Bruce F. Johnston, Food Research Institute, Stanford University: OECD Development Centre, Paris, summer 1970.

Donald Katzner, University of Pennsylvania: assistant professor, department of economics, University of Waterloo, 1970–71.

Mohamad W. Khouja, Oklahoma State University: economic consultant, Kuwait Fund for Economic Development, Middle East.

Alan P. Kirman, Johns Hopkins University: C.O.R.E., University of Louvain, Belgium. 1970–71.

J. William Leasure, San Diego State College: visiting Fulbright lecturer, Pedro Ruiz Gallo University, Peru, July 1970–Jan. 1971.

Hartley V. Lewis, University of British Columbia: research staff, Prices and Incomes Commission, Ottawa.

Clifton W. Loomis, Cornell University: advisor to the President of the Agricultural Development Fund of Iran.

Michael J. McCarthy, University of Pennsylvania: Wharton Economic Forecasting Associates, Inc., 1970–71.

Paul I. Mandell, Food Research Institute, Stanford University: director, Stanford Latin American Studies Undergraduate Program, Mexico City, fall and winter 1970–71.

Stanley H. Masters, Rutgers College: visiting research associate, Institute for Research on Poverty, University of Wisconsin.

Richard Meyer, Ohio State University: research and economic development, Sao Paulo, Brazil, 1970–72.

Edward H. Moscovitch, Williams College: Office of the Mayor, City of Boston.

Gordon R. Munro, University of British Columbia: Penang University, Malaysia.

Thomas X. O'Brien, Brandeis University: White House Fellow.

Michael Parti, Rutgers College: Center for Advanced Study in Behavioral Sciences, Stanford University.

E. C. Pasour, Jr., North Carolina State University: University of Chicago, 1970–71.

George F. Patrick, Purdue University: project specialist, Ford Foundation, Instituto de Pesquisas Economicas Aplicadas, Ministry of Planning, Brazil, Jan. 1970.

Jon A. Rusmussen, Wesleyan University: The Brookings Institution, Economic Policy Fellowship; Federal Water Pollution Control Administration.

Jati K. Sengupta, Iowa State University: visiting professor, Administrative Staff College of India, May 1970– Sept. 1970.

James F. Shepherd, University of California, Berkeley: visiting associate professor, department of economics, University of British Columbia, 1971.

William Silber, New York University: Council of Economic Advisors, 1970–71.

Michael K. Taussig, Rutgers College: visiting research associate, economics department, Princeton University.

Peter Timmer, Food Research Institute, Stanford University: associate advisor to the Indonesian Government with the Development Advisory Service of Harvard University, May 1970– Sept. 1971.

Stephen Turnovsky, University of Pennsylvania: assistant professor, department of economics, University of Toronto, 1970–71.

Robert P. Vichas, West Georgia College: Fulbright professor of economics, Jose Simeon Canas University, San Salvador, El Salfador, July 1- Dec. 31, 1970.

Terence B. Wales, University of Pennsylvania: visiting associate professor, department of economics, University of British Columbia, 1970–71.

Robert Willis, Wesleyan University: Research Fellow, National Bureau of Economic Research.

Gordon C. Winston, Williams College: economic advisor, Pakistan Institute of Development Economics, Karachi, Pakistan.

Resignations

Noel W. Allen, Auburn University.

W. Troy Anders, Auburn University: Shorter Junior College.

Terryl J. Anderson, College of Business Administration, University of Oregon, June 1970.

Richard A. Bilas, Georgia State University: California State College.

Richard G. Blackhurst, Rutgers College: Bologna Center of the School of Advanced International Studies, Johns Hopkins, June 1970.

Eric Brucker, Southern Illinois University: University of Delaware, June 1970.

Hector Correa, Tulane University, June 1970.

Coldwell Daniel III, University of Houston, June 1970.

John H. Duchman, University of Tennessee at Chattanooga.

Richard M. Furr, School of Business Administration, The American University, June 1970.

Ronald B. Gold, Ohio State University, June 1970.

Mark R. Greene, College of Business Administration, University of Oregon, Oct. 1970.

Kenneth R. Grubbs, University of Southern Mississippi: Louisiana State University.

Kanji Haitani, Southern Illinois University: State University of New York College at Fredonia.

James Heilbrun, Columbia University: Fordham University.

William R. Henry, North Carolina State University: Georgia State University, Aug. 1970.

George Hoffer, University of Virginia: Virginia Commonwealth University.

Udo E. Kramer, School of Business Administration, The American University, June 1970.

Roger C. Lawrence, Columbia University: UNCTAD, summer 1970.

Donald L. Losman, University of Tennessee at Chattanooga.

H. Nelson Lunn, School of Business Administration, University of North Carolina at Chapel Hill.

Spyros G. Makridakis, Rutgers College, June 1970.

Mary Malchow, University of Southern Mississippi.

H. O. Nourse, University of Illinois: University of Missouri-Saint Louis.

Dennis J. O'Connor, Arizona State University.

Leroy Pagano, School of Business Administration, The American University, June 1970.

John F. Pearce, Clemson University: North Georgia College, Sept. 1970.

Mark Powers, South Dakota State University: vice-president, research, Chicago Mercantile Exchange, Jan. 1970.

David J. Reid, Johns Hopkins University: London School of Economics, June 1970.

Brian S. Rungeling, Auburn University: University of Mississippi.

Richard H. Rush, School of Business Administration, The American University, June 1970.

Donald C. Snyder, Illinois Wesleyan University.

Peter E. Stangl, Wheaton College, July 1970.

Gordon R. Tush, University of Nebraska.

Donald L. Tuttle, School of Business Administratio University of North Carolina at Chapel Hill.

Wayne Vroman, Oberlin College.

Richard S. Wallace, Georgia State University School of Business Administration, Winthrop College

Richard H. Wood, Jr., University of Portland: Stetson University, Sept. 1970.

Robert Woodfin, Auburn University.

The American Economic Review

ARTICLES

DECEMBER 1970

THE AMERICAN ECONOMIC ASSOCIATI(

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September and December.

• Membership dues of the Association are $10.00 a year, $8.00 of each year's dues being in payment of a year's subscription both to the *Review* and the *Journal of Economic Literature*. Subscriptions by nonmembers are $10.00 a year. Single copies of the *Review*, $2.00; of the *Journal*, $2.00.

• Correspondence relating to the *Paper and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

HAROLD F. WILLIAMSON

Secretary-Treasurer of the American Economic Association

1962–70

Harold F. Williamson

Income Taxes and Housing

By Henry Aaron*

This paper presents new estimates of the distributional impact of special income tax provisions relating to housing. It also discusses possible secondary effects of these tax factors resulting from adjustments in the returns to capital, shifts in investment, and changes in the housing stock. The tax benefits to homeowners are shown to be equivalent to a reduction in the price of housing to homeowners that results in a benefit to them approximately equal to the tax savings.[1]

Several previous studies have approached this subject in a somewhat different manner from that employed here. Richard Goode estimated the revenue loss in 1958 from favorable tax treatment of homeowners on the basis of the 1959 Federal Reserve Board's *Survey of Consumer Finances* and other sources, but data then available permitted only global estimates of revenue effects. Melvin and Anne White examined the reduction in the variance in effective tax rates within income brackets arising from removal of favorable tax treatment of homeowners. David Laidler estimated price and income elasticities of demand for housing and applied these to estimate the welfare loss of present tax treatment of homeowners. Richard Slitor examined in detail tax treatment of housing (not just of owner occupants) and evaluated various proposals to encourage additions to and renovation of the housing stock. Robert Tinney, under the direction of David and Attiat Ott, employed procedures similar to those used here for the year 1964, but concentrated on welfare effects. Furthermore, that study dealt with adjusted gross income classes, rather than broad income classes used here, making comparison of results difficult.

I

The personal income tax encourages taxpayers to hold larger stocks of all consumer goods, including housing, than they would if imputed income from these goods were taxed as is money income from investment goods. Part of each household's portfolio consists of an inventory of durable commodity assets, such as houses, cars, furniture, and household appliances, which provide consumer services. Moreover, since the distinction between durables and nondurables is purely notional, the list, in principle, should include the household's stock of nondurables, clear down the scale of durability to clothing and food on the shelf. At each point in time, the household holds a stock of such assets with a money value $A(t)$. The stock changes continuously, by the amount of $dA(t)/dt$. If the household is in equilibrium, it receives a risk adjusted net rate of return, r, equal to the rate it receives from other assets as well.[2] That being the case, between two points in time the household

* The Brookings Institution and the University of Maryland. The author wishes to thank an anonymous referee, Kul Bhatia, Wilfred Lewis, and Joseph Pechman for helpful comments, Miss Slawka Pylyshenko for programming assistance, Mrs. Alicia Munnell for assistance in estimating imputed rent, Miss E. Morely Marshall for general assistance, and The Brookings Institution for supporting the study of the distribution of benefits of *U.S.* housing policies of which this paper is a part. The views expressed are those of the author and do not purport to represent the views of the other staff members, officers, or trustees of The Brookings Institution.

[1] Alternatively, the tax benefits may be regarded as a cash payment conditioned on housing purchases. If such a payment is taxable and it is stipulated that the payment must be large enough to leave the homeowner as well off as he is under current favorable tax treatment, the payment may exceed gross housing cost.

[2] As M. and A. White note (p. 231), the assumption of equilibrium is unrealistic, particularly with respect to housing. For lack of data, however, it is retained here.

receives imputed income, I, shown by equation (1).

$$(1) \qquad I = rA(t_1) + r\int_{t_1}^{t_2} dA(t)/dt\,dt$$

For income tax purposes, no imputations are made, thus encouraging households to keep more of their assets in forms which yield imputed rather than cash income. The only imputation on consumer assets included in national income statistics is rent on owner occupied houses.

The amount of imputed income on owner occupied houses equals the net income the homeowner could obtain by renting his residence to another household or by investing his equity interest in the house in some other capital asset; he thereby avoids rental expenses he would have to incur if he lived in a rented unit. For this reason, a homeowner will have less cash income and fewer cash expenditures than a renter with identical earnings and total assets. This situation is depicted in Table 1.[3] If the renter and owner earn the same incomes from work and assets and pay the same price for housing, residual income before tax will be the same. Under a tax system which treated income from all sources alike, tax liabilities on these two households would be identical and disposable income for consumption of goods other than housing or for saving would be the same. Under the *U.S.* personal income tax, the difference is substantial.

The item which should be imputed as income, net rent *NR*, is equal to gross rent *GR* less maintenance expense M, depreciation D, mortgage interest I, and property and other indirect taxes T levied on the homeowner. If the homeowner were treated like other businesses, he would be required to report *GR* as income and he would be

[3] Table 1 is similar to the example presented by Goode.

TABLE 1—ILLUSTRATIONS OF TAX TREATMENT OF OWNERS AND RENTERS

	Renter	Owner
Earnings	\$10,000	\$10,000
Income from assets at 4 Percent		
Assets valued at \$25,000	1,000	—
Assets valued at \$15,000	—	600
Equity of \$10,000 in house	—	400
Money income	11,000	10,600
Rent payments	2,500	—
Housing expenses		2,100
Residual money income	8,500	8,500
Tax liability[a]	1,304	962

[a] Tax liability is computed under the following assumptions: a) real housing costs are 25 percent of earnings for both renter and owner, including \$400 net profit or net imputed rent, \$900 in mortgage interest (6 percent on \$15,000 mortgage) and property taxes of \$500; b) both renter and owner are four-person households with no aged members; c) renter claims standard deduction; d) homeowner itemizes and claims as deductions, \$1,400 in mortgage interest and property taxes, and \$1,000 in other deductions.

permitted business deductions in the amount $M+D+I+T$, generating income, positive or negative, of *NR* on which he would be required to pay tax. In fact, the homeowner is not required to report *GR*, but he is allowed deductions in the amount $I+T$. Current tax treatment thus understates taxable income relative to other asset holders by $NR+I+T$.

This section reports estimates of the revenue effect of the failure to tax imputed rent on owner occupied residences while permitting owners deductions for property taxes and mortgage interest paid. The tabulations presented in this section compare tax liabilities under three alternative laws with those under present code provisions:

- Law 1 Disallow deductions for both property taxes and mortgage interest.
- Law 2 Include net imputed rent in taxable income.
- Law 3 Include net imputed rent in taxable income and disallow deductions for both property taxes and mortgage interest.

Law 3 would place homeowners, qua owners of capital, on an equal tax footing with owners of other assets and it would place owners, qua tenants, on an equal footing with renters.

The effects of disallowing the deductions were simulated on the Brookings Tax File for the tax year 1966 and the tax law then in effect. Imputed rent was estimated from net equity in houses by income class as reported in G. Katona et al., and D. Projector and G. Weiss.[4] Imputed rent was estimated on the alternate assumptions of a 4 percent and a 6 percent yield on the owner's equity interest in his house. Even the 4 percent yield implies imputed net rent far in excess of the amounts reported in the national income accounts ($21.2 billion versus $12.1 billion in 1966). Yields on long-term government bonds averaged 5.16 percent, those on high grade corporate securities averaged 5.13 percent, and those on FHA mortgages averaged 6.29 percent throughout 1966 and each has risen since then. Yields on high grade municipal bonds averaged 3.82 percent. Since income from home ownership is (more than) exempt from tax, the tabulations presented here embody the 4 percent assumption.[5] Readers who find the 4 percent estimate implausible may adjust the estimates accordingly.

As shown in Table 2, homeowners pay $7.0 to $9.0 billion less in taxes under current law than they would under Law 3 in the absence of quantity adjustments.[6] This amount is 16.6 to 21.5 percent of the $42.0 billion collected from homeowners under the personal income tax.[7] More detailed results are presented below.

TABLE 2—CHANGE IN REVENUES: THREE LAWS COMPARED WITH PRESENT INCOME TAX PROVISIONS

	Change in Tax (billions) 4 percent	6 percent
Law 1	2.9	2.9
Law 2	4.0	6.0
Law 3	7.0[a]	9.0[a]

[a] The revenue effect of Law 3 exceeds the sum of the effects of Laws 1 and 2 separately because some taxpayers become subject to higher marginal brackets when all changes are made simultaneously than when each change is made independently.

Tax liabilities differ sharply under current tax law within each income bracket among various socioeconomic groups. Many aspects of the tax code contribute to these variations. Favorable treatment of homeowners is one of these reasons.[8] The effect of all these factors on selected subgroups of taxpayers is shown in Table 3. As is apparent, homeowners tend to pay less tax at each bracket than do non-homeowners. The aged pay less than the non-aged. Taxpayers with sufficient deductions to itemize tend to pay less than those who claim the standard deduction.

Increases in tax liabilities and in average effective tax rates are shown in Tables 4, 5, and 6. In evaluating these estimates, the reader should keep in mind that adoption of any of the alternate tax laws would require other fiscal or monetary adjustments—lowered taxes, increased expenditures, or lowered interest rates—to prevent

[4] The procedures used in estimating imputed rent are described in the Appendix.

[5] The assumption of a 6 percent rate of return is made and defended by Laidler.

[6] Here and elsewhere in this article, revenue effects are calculated as if the pattern of homeownership and the housing stock would be the same under the alternative laws as they are under current law. Quite obviously reactions to any large change in tax liability would be substantial, but for reasons indicated below there is considerable doubt regarding the ultimate magnitude of these reactions. The text estimates indicate with some accuracy the revenue effects in the first year after adoption of the indicated tax provisions, not the subsequent effects.

[7] All tax statistics are computed from the Brookings Tax File and are subject to sampling error. For example, total estimated tax collections for 1966 are $56.5 billion. Actual collections were $56.1 billion, according to *Statistics of Income, Individual Income Tax Returns*, 1966.

[8] For the best treatment of the proportion of such variation in tax liability within income brackets due to special tax treatment of homeowners, see M. and A. White.

TABLE 3—FEDERAL INCOME TAX COLLECTIONS UNDER CURRENT LAW, AS PERCENT OF TOTAL INCOME[a]

	Total Income Class									
	$0–3,000	$3,000–5,000	$5,000–7,000	$7,000–10,000	$10,000–15,000	$15,000–25,000	$25,000–50,000	$50,000–100,000	$100,000–9,999,999	All Classes
All Returns	3.7	6.7	8.0	8.9	10.4	12.7	17.1	24.5	28.9	11.0
Homeowners	2.5	4.9	7.1	8.2	9.9	12.2	16.9	24.4	28.9	11.3
(1) Non-aged	2.8	5.5	7.3	8.2	9.9	12.3	17.0	24.6	28.6	11.1
(a) Itemized	7.7	4.6	5.8	7.4	9.4	12.0	16.7	24.5	28.6	11.6
(b) Standard Deduction	2.4	6.1	8.5	9.7	11.7	14.2	19.8	27.0	27.8	9.8
(2) Aged	0.8	2.5	5.4	7.5	9.8	11.8	16.4	23.3	29.4	13.5
(a) Itemized	1.1	2.3	4.9	6.8	9.3	11.5	16.5	23.4	29.5	14.6
(b) Standard Deduction	0.6	2.9	7.2	10.1	11.2	12.8	16.4	20.5	23.2	8.1
Non-Homeowners	4.2	7.8	9.3	10.5	12.5	15.5	18.4	25.4	28.6	10.1
(1) Non-aged	4.7	8.0	9.4	10.5	12.5	15.6	18.4	26.6	27.8	10.2
(a) Itemized	5.0	7.0	8.7	9.3	11.6	15.0	18.9	26.4	28.2	11.0
(b) Standard Deduction	4.7	8.2	9.9	11.1	13.1	16.1	16.3	30.8	16.3	9.8
(2) Aged	1.0	5.0	8.2	9.6	11.7	14.9	18.5	23.2	29.8	9.5
(a) Itemized	1.6	4.5	7.6	8.4	11.9	14.3	18.2	23.0	29.8	12.1
(b) Standard Deduction	0.9	5.3	9.0	10.5	11.3	16.0	21.2	30.2	27.4	6.8
All Aged	1.4	2.9	5.7	7.0	9.9	12.1	16.8	23.3	29.5	12.4
All Nonaged	4.1	7.2	8.2	8.9	10.4	12.7	17.1	24.8	28.6	10.8

Source: See Appendix.

[a] Total income equals adjusted gross income plus excluded dividends, excluded sick pay and imputed rent. Capital gains and losses are included in income in full in the year realized but no losses carried forward from previous years are deducted. Total income does *not* include unrealized capital gains, tax exempt interest, or the excess of percentage over cost depletion on income from natural resources.

TABLE 4—CHANGE IN TAX COLLECTIONS PER FAMILY FROM DISALLOWANCE OF DEDUCTIONS FOR MORTGAGE INTEREST AND PROPERTY TAXES AND CHANGE IN TAX AS PERCENT OF TOTAL INCOME BY TOTAL INCOME CLASS[a]

$ = Average change in tax from current law per family in dollars.
% = Average change in tax as percent of average total income.

	Total Income Class									Total Tax Savings
	$0–3,000	$3,000–5,000	$5,000–7,000	$7,000–10,000	$10,000–15,000	$15,000–25,000	$25,000–50,000	$50,000–100,000	$100,000–9,999,999	(millions of $)
All Returns										
$	1	8	19	45	85	166	292	517	1,082	2,904
%	0.1	0.2	0.3	0.5	0.7	0.9	0.9	0.8	0.5	
Homeowners										
$	3	21	32	64	106	196	329	562	1,144	2,904
%	0.2	0.5	0.5	0.8	0.9	1.1	1.0	0.9	0.5	
Non-aged Itemized										
$	32	52	68	99	139	225	368	581	1,087	2,685
%	2.4	1.3	1.1	1.2	1.1	1.2	1.1	0.9	0.5	
Aged Itemized										
$	17	35	59	77	97	162	270	600	1,360	219
%	0.8	0.9	1.0	0.9	0.8	0.9	0.8	0.9	0.6	

Source: See Appendix.

[a] See note a, Table 3.

TABLE 5—CHANGE IN TAX COLLECTIONS PER FAMILY FROM INCLUSION IN ADJUSTED GROSS INCOME OF IMPUTED NET RENT ON OWNER OCCUPIED HOUSES, AND CHANGE IN TAX AS PERCENT OF TOTAL INCOME, BY TOTAL INCOME CLASS[a]

\$ = Average change in tax from current law per family in dollars.
% = Average change in tax as percent of average total income.

	Total Income Class									Total Tax Savings
	\$0–3,000	\$3,000–5,000	\$5,000–7,000	\$7,000–10,000	\$10,000–15,000	\$15,000–25,000	\$25,000–50,000	\$50,000–100,000	\$100,000–9,999,999	(millions of \$)
All Returns										
\$	11	20	36	58	103	180	433	654	1,217	4,017
%	0.8	0.5	0.6	0.7	0.9	1.0	1.3	1.0	0.5	
Homeowners										
\$	36	54	60	83	128	212	488	712	1,286	4,017
%	2.4	1.4	1.0	1.0	1.1	1.2	1.4	1.1	0.6	
Non-aged										
\$	38	56	61	82	128	213	481	678	1,211	3,659
%	2.6	1.4	1.0	1.0	1.1	1.2	1.4	1.0	0.6	
Itemized										
\$	42	49	57	80	126	210	482	679	1,219	2,543
%	3.2	1.2	0.9	0.9	1.0	1.1	1.4	1.0	0.6	
Standard Deduction										
\$	38	60	64	87	136	239	462	644	919	1,115
%	2.5	1.5	1.1	1.0	1.1	1.3	1.3	1.0	0.5	
Aged										
\$	23	47	57	88	133	201	529	851	1,459	358
%	1.3	1.2	1.0	1.1	1.1	1.1	1.6	1.3	0.6	
Itemized										
\$	26	40	57	88	132	195	525	856	1,466	287
%	1.2	1.0	1.0	1.1	1.1	1.0	1.5	1.3	0.6	
Standard Deduction										
\$	22	62	57	87	136	223	589	697	1,055	71
%	1.3	1.6	0.9	1.0	1.1	1.2	1.8	1.0	0.6	

Source: See Appendix.
[a] See note a, Table 3.

unemployment. Each such compensating policy would have distributional consequences of its own. Strictly speaking, the distributional consequences of the alternative tax laws described here should be combined with those of whichever compensating policy is adopted. This procedure is not followed here since the purpose of this paper is to examine the non-neutrality of the current tax system with respect to housing. In pursuing this goal, consideration of the extremely diverse possible patterns of benefits from compensating fiscal or monetary actions would obscure, rather than clarify, matters.

Deductibility of mortgage interest and property taxes reduces revenues by \$2.9 billion (see Table 4).[9] This saving accrues entirely to homeowners who itemize. As a fraction of income, tax savings for all homeowners rise until annual income exceeds \$25,000. Deductibility of interest is worth less to the aged at each income bracket because more aged homeowners have paid off a larger part of their mortgage debt than have non-aged taxpayers.

[9] Deductibility of mortgage interest alone reduces revenues by \$1.6 billion. Deductibility of property taxes alone reduces revenues by \$1.5 billion. The revenue effects of disallowing deductions for both mortgage interest and property taxes is less than the sum of disallowing each separately because the standard deduction becomes effective for more tax payers when both are disallowed than when either is disallowed separately.

TABLE 6—CHANGE IN TAX COLLECTIONS PER FAMILY FROM INCLUSION IN ADJUSTED GROSS INCOME OF IMPUTED NET RENT ON OWNER OCCUPIED HOUSES AND DISALLOWANCE OF DEDUCTIONS FOR MORTGAGE INTEREST AND PROPERTY TAXES, CHANGE IN TAX AS PERCENT OF INCOME, AND TAX AS PERCENT OF INCOME, BY TOTAL INCOME CLASS[a]

$=Average change in tax from current law per family in dollars.
%=Average change in tax as percent of average total income.
T=Tax rate under reformed law.

	Total Income Class									Total Tax Savings
	$0–3,000	$3,000–5,000	$5,000–7,000	$7,000–10,000	$10,000–15,000	$15,000–25,000	$25,000–50,000	$50,000–100,000	$100,000–9,999,999	(millions of $)
All Returns										
$	12	29	56	103	190	351	736	1,179	2,308	6,982
%	0.9	0.7	0.9	1.2	1.6	1.9	2.2	1.8	1.0	
T	4.7	7.6	9.1	10.3	12.3	15.0	19.8	26.7	30.0	
Homeowners										
$	40	77	93	146	237	414	830	1,283	2,440	6,982
%	2.6	1.9	1.5	1.7	2.0	2.3	2.5	1.9	1.1	
T	5.5	7.4	9.0	10.2	12.3	14.9	19.9	26.8	30.1	
Non-aged										
$	42	77	92	146	238	420	836	1,244	2,282	6,394
%	2.8	1.9	1.5	1.7	2.0	2.3	2.5	1.9	1.0	
T	6.1	7.9	9.1	10.2	12.3	15.0	20.0	27.0	29.8	
Itemized										
$	82	105	127	178	267	442	865	1,269	2,316	5,279
%	6.2	2.6	2.1	2.1	2.2	2.4	2.6	1.9	1.1	
T	13.0	7.5	8.1	9.7	12.0	14.9	19.9	26.9	29.8	
Standard Deduction										
$	38	60	64	87	136	239	462	644	919	1,115
%	2.5	1.5	1.1	1.0	1.1	1.3	1.3	1.0	0.5	
T	5.5	8.2	9.9	11.2	13.3	16.1	21.9	29.0	28.1	
Aged										
$	31	77	105	148	210	333	790	1,441	2,804	588
%	1.7	2.0	1.8	1.8	1.7	1.8	2.3	2.1	1.2	
T	2.6	5.0	7.7	9.6	11.9	13.9	19.3	25.7	30.8	
Itemized										
$	55	82	117	166	231	362	804	1,465	2,832	517
%	2.6	2.1	2.0	2.0	1.9	1.9	2.3	2.2	1.2	
T	3.7	4.9	7.4	9.0	11.6	13.8	19.4	25.9	30.9	
Standard Deduction										
$	22	62	57	87	136	223	589	697	1,055	71
%	1.3	1.6	0.9	1.0	1.1	1.2	1.8	1.0	0.6	
T	2.0	5.4	8.9	11.5	12.8	14.3	19.4	21.4	24.4	

Source: See Appendix.
[a] See Note a, Table 3.

The failure to tax imputed net rent generates tax savings worth even more—$4.0 billion—than the two deductibility provisions combined (see Table 5). The nontaxation of imputed rent helps homeowners who claim the standard deduction as well as those who itemize, since the inclusion of net imputed rent would raise a household's adjusted gross income and taxable income whether or not it itemized deductions. The estimates shown in Table 5 do not include present nonfilers who would be required to file returns and pay taxes if imputed rent were taxable. For this reason they understate the revenue effects. As a fraction of income, tax benefits for nontaxability of imputed rent are highest for taxpayers in the lowest bracket largely because these

TABLE 7—DIFFERENCE IN TAX RATES AS PERCENT OF TOTAL INCOME BETWEEN HOMEOWNERS AND NON-HOMEOWNERS—CURRENT LAW AND LAW 3[a]

	Total Income Class								
	\$0–3,000	\$3,000–5,000	\$5,000–7,000	\$7,000–10,000	\$10,000–15,000	\$15,000–25,000	\$25,000–50,000	\$50,000–100,000	\$100,000–9,999,999
Non-aged									
Current Law	1.5	2.0	1.9	2.0	2.2	2.9	0.9	1.4	−0.9
Law 3	−1.3	0.1	0.3	0.3	0.3	0.6	−1.6	−0.5	−2.0
Itemized									
Current Law	−1.8	2.1	2.7	1.7	1.8	2.5	1.6	1.4	−0.6
Law 3	−8.0	−0.5	0.6	−0.4	−0.4	0.1	−0.9	−0.5	−1.7
Standard Deduction									
Current Law	1.8	1.5	1.0	0.9	1.0	1.4	−4.3	2.8	−11.3
Law 3	−0.8	0.0	−0.1	−0.1	−0.1	0.0	−5.6	1.8	−11.8
Aged									
Current Law	0.2	1.9	2.3	1.8	1.5	2.8	1.5	−0.4	0.1
Law 3	−1.5	0.0	0.5	0.0	−0.2	1.0	−0.8	−2.5	−1.0
Itemized									
Current Law	0.5	1.8	2.2	1.4	2.2	2.4	1.2	−0.7	0.1
Law 3	−2.1	−0.3	0.2	−0.6	0.3	0.5	−1.1	−2.9	−1.1
Standard Deduction									
Current Law	0.1	1.5	1.0	0.0	−0.4	2.9	3.6	9.7	3.7
Law 3	−1.1	−0.1	0.1	−1.0	−1.5	1.7	1.8	8.7	3.1
All									
Current Law	1.4	2.4	1.9	2.0	2.2	2.8	1.0	0.6	−0.5
Law 3	−1.3	0.4	0.4	0.3	0.2	0.5	−1.5	−1.3	−1.6

Source: See Appendix.

[a] Minus sign indicates that taxes of homeowners as a percent of total money income are greater than those of non-homeowners.

tax brackets include many households with temporarily depressed incomes. The fraction rises over the income range of \$5,000 to \$50,000.

The revenue effects of the full taxation of income from homeownership are shown in Table 6. On balance, such a change in the tax laws would narrow the difference in tax treatment between homeowners and other taxpayers (see Table 7).[10] Such narrowing occurs for virtually all taxpayers in the tax brackets between \$3,000 and \$25,000.[11] In the lowest and highest income brackets the pattern becomes more diverse, partly because of sampling variability and partly because of the concentration in top and bottom brackets of households for whom loss carry forward and capital gains crucially influence effective rax rates.[12]

II

The foregoing tabulations demonstrate who, in an accounting sense, receives benefits from special income tax treatment of housing. Though such patterns are interesting, they may differ substantially from the distribution of benefits which would have prevailed before the market responses to the tax benefits have worked themselves out. The size of these responses will depend on the size of tax benefits relative to housing expenditures and on household responses to these price effects. Fur-

[10] These results are entirely consistent with, and indeed, are implied by Melvin and Anne White.

[11] The only exceptions are aged households with incomes between \$7,000 and \$15,000 who do not itemize.

[12] In a number of instances, the reformed tax treatment of homeowners would cause them to be taxed more heavily than are nonhomeowners. This situation will arise, apart from sampling variability, if homeowners receive less of their current income in the form of realized net long-term capital gains or have smaller loss carryforwards than do nonhomeowners.

thermore, these tabulations do not indicate any changes in factor incomes that may have occurred during the adjustment process. Upon the resolution of these issues depends the ultimate incidence of tax benefits to homeowners. As will be shown, available data are inadequate to resolve these issues but do suggest some probable consequences. While hard information on the composition of gross rent is unavailable, it is possible to indicate the rough order of magnitude of the tax subsidy.

Market Effects of Homeowner Tax Subsidies

The tax benefits to owner occupants affect homeowners in either or both of two ways. As consumers, they find that homeownership is cheaper relative to rental housing (and other consumer goods) than it would be if income from housing were fully taxed. As investors, they find that the after-tax investment income from an owner occupied home relative to after-tax income from other investments is greater than it would be if income from housing were fully taxed.

As consumers, households can respond to tax benefits in either of two ways. Some households, who would have rented in the absence of tax benefits, will choose to own their residence. Because of the tax benefits, some households, who would have owned even without tax benefits, will choose to live in a more valuable residence than the residence they would otherwise have selected. There is no direct evidence on the extent to which the tax differentials affect household choice between owning and renting.

If this consumer choice is insensitive to the relative price of rented and owned units, then favorable tax treatment of owner occupants will cause few households to become owners rather than tenants. In the extreme case, if the cross price elasticity of demand between housing to rent and housing to own were zero, demand for rented units and hence construction of rental units would be unaffected by tax benefits to homeowners. Homeowners might purchase larger or more luxurious houses than they would have done in the absence of tax benefits. In addition, they might purchase more of other consumer goods or save more.

At the other extreme, if the cross elasticity were infinite, tax benefits would cause all households to prefer to become owners.[13]

To the extent that favorable tax treatment of homeowners causes demand for housing units "to buy" to rise at the expense of housing units "to rent," the following effects will occur: 1) rents will tend to decline; 2) prices of housing units for sale to owner occupants will tend to rise; 3) housing units, previously rented, will be sold to owner occupants; 4) construction of new rental units will fall and that of units for sale to owner occupants will rise. The greater the sensitivity of household choices between forms of tenure, the greater will be effects 1) and 2). These consumption reactions, in turn, will generate changes in the housing stock through effects 3) and 4).

If construction costs are unaffected by the composition of construction activity, then, over time, the fraction of the housing stock which is owner occupied will tend to rise as a result of favorable tax treatment of owner occupants. This situation is depicted in Figure 1.[14] However, there are

[13] Note that the distinction between owning and renting involves differences in responsibilities and risks in addition to, though perhaps resulting from, "mere" differences in legal tenure. For this reason the package of residential services purchased by an owner differs from that purchased by a renter. For a contrary opinion see R. Muth 1969 (p. 19).

[14] In Figures 1 and 2, quantity of housing refers to a composite bundle of services. Different housing units in general will embody different quantities of housing since they contribute to the production of different numbers of units of such housing services. This distinction between structures and services is also used in Muth 1969, (pp. 18–19).

reasons to expect capital costs to be affected, as argued below. Housing units vary in price. The quality or "amount" of housing owners want to own, as well as the choice between owning and renting, depends on the relative price of housing to own and housing to rent. Accordingly, the quantity of housing is measured in units of, say, \$1,000, measured in initial prices.

As with excise taxes and subsidies, tax benefits may be represented diagrammatically as either a shift in seller perceived demand, the supply curve fixed, or as a shift in buyer perceived supply, the demand curve fixed. Figure 1a follows the latter course. Tax benefits of b[15] shift the supply curve from S_0 to S_1. As a result the quantity of housing services demanded for owner occupancy would rise from q_0 to q_1. The rental demand for housing services would shift downward, reducing the demanded amount of housing for rent from q_0' to q_1'. Until the stock of rental properties declined to the equilibrium demanded, q_1', renters might derive benefits from the introduction of favorable tax treatment for homeowners. Households which would have been owner occupants in the absence of tax benefits and whose housing consumption is unchanged, enjoy a subsidy in the amount b times the quantity of housing consumed. Households which are induced to become homeowners because of the tax benefits, or to increase their consumption of housing, enjoy a subsidy to the extent that the tax subsidy exceeds the amount necessary to cause the household to become a homeowner. For example, the subsidy is AB in Figure 1a times the quantity of housing consumed for the household which would become a homeowner when gross rent less tax benefits was BC. The amount DB, which is in excess of the amount that is just necessary to induce the household to own, is dead weight loss unless homeownership is thought to provide social benefits. The area EFG measures the dead weight loss (offset by social benefits, if there are any), from causing homeownership to rise from q_0 to q_1.

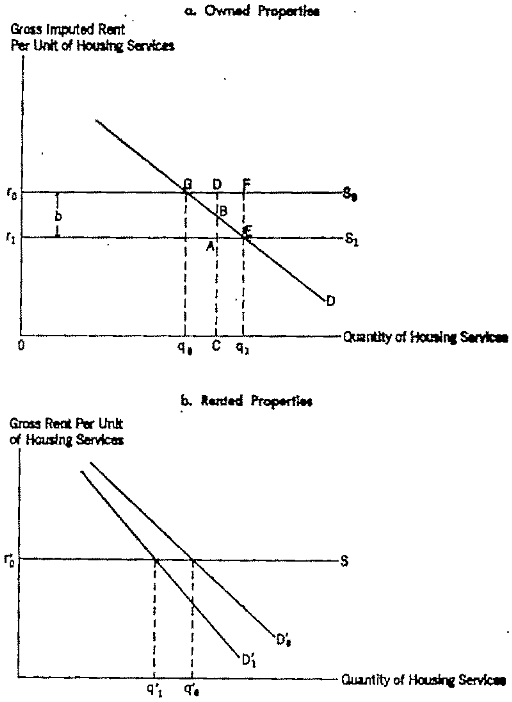


FIGURE 1. IMPACT OF FAVORABLE TAX TREATMENT OF HOMEOWNERS

As drawn, Figure 1 indicates a rise in consumption of owner occupied housing greater than the decline in consumption of rental housing. In addition, the gross unit cost of housing is assumed not to rise with output. Since tax benefits to homeowners reduce the net price of housing on the average, an increase in demand for housing in general is plausible.[16] However, a shift of capital to housing from other industries can be accomplished only if the return to capital invested in housing rises.[17]

[15] The size of b in relation to r_0 depends on the taxpayers marginal rate as well as other tax provisions. See Table 8.

[16] The analysis in the following paragraphs is modeled on the article by Mieszkowski.

[17] The total flow of savings is assumed to be unaffected by tax benefits to homeowners, thus precluding a

Since capital can move to whatever use yields the highest rate of return net of risk, such a shift of capital can occur only if the rate of return to capital in general rises.[18] This result follows from the fact that the production of housing services is more capital intensive than the production of nearly all other commodities.[19]

As a result of this rise in the price of capital and of the capital intensity of the production of housing services, the price of *all* housing services will rise somewhat relative to the price of other commodities and the factor income of *all* wealth holders will rise somewhat relative to all non-wealth holders. Tax benefits to homeowners will thus have three distinct distributional effects. First, direct recipients of tax benefits will receive a subsidy in the amount of the benefits. Second, all consumers of housing services will suffer a loss due to the rise in the price of housing services which, for homeowners, will partially offset the first effect. Third, all recipients of income from capital will benefit from an increase in the rate of return to capital. Some households will experience more than one of these effects.

The size of each of these effects depends on a number of factors. In general, the first effect will be larger the more sensitive is demand to changes in the relative prices of owned homes, on the one hand, and of rental housing and other commodities on the other. The second and third effects will be larger the larger the difference between the housing industry and other industries in capital intensity and the smaller is the elasticity of substitution between capital and labor in housing and in other industries. This relationship exists because large disparities in factor intensities and small values of the elasticities of substitution require a large rise in the price of capital to restore equilibrium after the introduction of a negative commodity tax in a capital intensive industry. The production of housing services is a far more capital intensive activity than is the production of most other goods and services. In the remainder of this section, however, the second and third effects will be ignored, not because they are unimportant but because exploration of these effects empirically would constitute a major separate study. It seems likely that the second and third effects in combination will work in the same general direction as the first, since increases in the price of capital would be distributed (at least) as regressively as are the benefits shown in Tables 4 through 6.

The magnitude of the shifts in the quantity of housing demanded per dollar of tax benefits depends in part on the elasticities underlying the demand schedules of Figure 1.

If demand were completely inelastic, no shift would occur. The proportion of the tax subsidy which accrues for homeownership expenditures that would have been made without tax benefits similarly depends on the number of renters who are induced by tax benefits to own, and the degree to which existing homeowners move to more costly residences because of tax benefits.

Some indication of the magnitude of the response in housing demand to tax benefits

perfectly elastic supply of capital to any industry. Thus the supply curve in Figure 1 is an abstraction which does not reflect changes in the return to capital.

[18] The text analysis presumes diminishing marginal returns to capital. If the supply of saving is unaffected, additional capital can be diverted to housing from other activities only by diverting it from uses where the yield on capital previously was higher than the yield in alternative uses, viz. housing. Only if the supply of saving were perfectly elastic with respect to the interest rate would an increase in capital in housing with no increase in yields be possible. See Mieszkowski, 1967.

[19] The extreme capital intensity of the production of housing services should not be confused with the lower capital-labor ratio in construction. See, for example, Edward Denison, "Changes in net output [of housing services] . . . are due almost entirely to changes in the quantity of capital and land used in the industry; labor input is almost negligible. . ." (p. 124).

can be obtained from estimates of the price elasticity of demand for housing—the percentage by which the quantity of housing demanded responds to a 1 percent change in prices. Major recent studies all suggest that the price elasticity of demand for housing is in the vicinity of -1 to -1.5, that is, a tax benefit which has the effect of reducing housing costs by 1 percent will raise the quantity of housing demanded by about 1 to 1.5 percent.[20]

These results suggest that if tax benefits for homeowners reduce gross rent, say, by 10–15 percent, they will increase demand for owner occupied housing by at least 10 percent and perhaps more than 20 percent. The accuracy of these estimates declines as one moves away from the price levels employed in the studies. Moreover, all of the cited studies, except that by Reid are based on time-series and contain no specific price elasticities for different socioeconomic groups.[21] The similarity of Reid's estimates with others offers only slight comfort that the estimated price elasticity of -1 applies to households in different income brackets.

The Size of Tax Benefits

Current income tax provisions understate income of homeowners by the amount $NR+I+T=GR-M-D$, and tax liability by $\tau(NR+I+T)=\tau(GR-M-D)$ where τ is the average tax rate on the excluded income. This reduction in tax liability is an implicit subsidy. The size of the subsidy as a fraction of gross rent is shown in Table 8. It depends positively on the fraction of gross rent absorbed by deductible expenses or net rent and on the marginal tax rate of the taxpayer.[22] The subsidy varies from negligible quantities for low bracket households who reside in houses with high maintenance costs and depreciation to very large amounts for high bracket households residing in properties with low maintenance and depreciation.

TABLE 8—SUBSIDY AS A PERCENT OF GROSS RENT

Imputed Rent Plus Deductible Expenses as a Fraction of Gross Rent	Tax Brackets (Percentage) 14	20	30	50	70
1/8	1.8	2.5	3.8	6.2	8.8
1/4	3.5	5.0	7.5	12.5	17.5
3/8	5.2	7.5	11.2	18.8	26.2
1/2	7.0	10.0	15.0	25.0	35.0
5/8	8.8	12.5	18.8	31.2	43.8
3/4	10.5	15.0	22.5	37.5	52.5
7/8	12.2	17.5	26.2	43.8	61.2
1	14.0	20.0	30.0	50.0	70.0

Little information is available on the composition of gross rents of owner occupied houses, however. A rule of thumb of the real estate market suggests that annual rents on single family houses should run 10–12 percent of market price.[23] Maintenance plus depreciation as a frac-

[20] See, for example, Tong Hun Lee (pp. 85–87), Margaret Reid, (p. 381), Richard Muth 1960, (pp. 72–75), H. Houthakker and L. Taylor (pp. 76–77). The dependent variables in these studies are: per capita net new construction of non-farm housing; (see Muth, 1960); real household expenditures on rent (see Reid); per family gross real non-farm residential construction (see Lee); per capita real rental value of residential construction costs-brick (see Lee, Muth 1960); the rent component of the consumers price index (see Reid); the implicit price deflator for imputed rent divided by the implicit price deflator for personal consumption expenditures (see Houthakker and Taylor).

[21] For a discussion of the probable importance of socioeconomic factors in housing demand, see R. Hartman.

[22] Alternatively, the tax saving from undertaxation of income from housing may be regarded as a cash tied subsidy, S, combined with full taxation of income from housing, which leaves the recipient as well off as he is under current law. If the subsidy itself is taxable, then $S(1-\tau)=\tau(GR\text{-}M\text{-}D)$ or $S=[\tau/(1-\tau)](GR\text{-}M\text{-}D)$. The size of such a subsidy may exceed gross rent.

[23] This rule evolved when interest rates were below current levels. Whether the trend toward higher interest rates would cause that rule to be altered is not clear. To the extent that the rise in interest rate is due to anticipated inflation rather than to an increase in the real rate of time preference, anticipated depreciation would tend to fall or anticipated appreciation would tend to increase as much as interest rates change.

TABLE 9—INCIDENCE OF HOMEOWNERSHIP AND RATIO OF DEBT TO ESTIMATED VALUE OF HOUSE, BY INCOME CLASS AND AGE GROUP

	Head Under 35		Head 35–54		Head 54–64		Head 65 and over		All Ages	
Income Class	Percent Home-owners	Percent Debt to Value[a]	Percent Home-owners	Percent Debt to Value[a]	Percent Home-owners	Percent Debt to Value[a]	Percent Home-owners	Percent Debt to Value[a]	Percent Home-owners	Percent Debt to Value[a]
Total	34	64.29	63	44.15	67	23.78	61	7.78	57	30.90
$0–$2,999	7	22.48	30	33.68	45	8.37	55	4.39	40	9.74
3,000– 4,999	18	71.68	45	41.85	68	13.65	68	8.42	45	41.77
5,000– 7,499	45	65.02	64	37.95	75	32.15	52	5.99	60	38.77
7,500– 9,999	58	66.60	75	39.41	86	13.77	85	7.92	73	37.80
10,000–14,999	50	75.06	86	37.17	88	20.72	72	15.06	80	35.51
15,000–24,999	94	56.64	87	40.47	78	16.43	91	1.81	86	34.38
25,000–49,999	62	16.10	92	30.65	93	11.92	94	22.14	92	24.42
50,000–99,999	100	13.76	98	21.13	96	37.41	95	1.47	94	11.82
100,000 and over			96	14.09	91	3.99	98	9.96	96	3.99

Source: [14, Table A-8 and A-14].
[a] Home value taken as sum of debt secured by owned home and estimated equity in owned home.

tion of market value has been estimated at from 2 to 6.5 percent.[24] This range suggests that tax favored elements, on the average, constitute at least one-third and as much as five-sixths of gross rent. Laidler (p. 61) estimates tax favored elements at 68 percent of gross rent. Because most land does not depreciate and much appreciates, and because the ratio of land value to property value is positively related to property value, it seems likely that the fraction of gross rent accounted for by tax favored items rises with income.[25] The plausible range of implicit subsidies runs from roughly 5 percent of gross rent in the lowest tax brackets to more than 50 percent in the highest brackets. Of course, the subsidy to households with no taxable income is zero.

If a household faces the choice of owning its residence or of renting it and investing the saved equity in other assets, the tax benefits make it increasingly likely, as one moves up the income scale, that ownership would be preferred and would seem to make a high debt-value ratio increasingly desirable. In fact, however, the ratio of mortgage debt to estimated value of owner occupied houses is negatively correlated with income, even within age cohorts (see Table 9).[26]

The tax subsidies shown in Table 8 raise substantially the consumption of housing by owner occupants. The change in housing demand attributable to the tax system rises with income unless the true price elasticity declines faster than the tax subsidy increases. For example, assume that the price elasticity demand is -2 for a household in the 14 percent bracket, three-eighths of whose gross rent consists of tax favored items. Existing tax subsidies would cause such a household to consume 10.4 percent more housing than it would consume in the absence of tax benefits. A household in the 50 percent bracket, five-eighths of whose gross rent consists of tax favored items, would ex-

[24] See J. Shelton (pp. 65, 67) for a survey of these estimates.

[25] On recent trends in land values, see A. Manvel.

[26] Two possible explanations for this paradox come to mind. First, lending institutions customarily will not lend as large a fraction of price on expensive as on inexpensive houses. Second, high income taxpayers are probably more likely than low income taxpayers to have inherited houses owned free and clear.

pand its housing consumption by more than 10.4 percent unless its price elasticity of demand were less than $-.33$. The conclusion that tax benefits to homeowners tend to raise consumption of housing seems incontrovertible and that it does so more for households in upper brackets than for those in lower brackets seems highly probable.

Tax Benefits on Rental Property

Whether the income tax system encourages homeownership relative to renting depends on the size and incidence of any tax benefits on rental housing. The major tax benefit not available to homeowners which may be available on rental property is deduction for depreciation in excess of true economic depreciation.[27] The bias of the tax system in favor of homeownership is offset to the extent that allowable depreciation for tax purposes on rented housing units exceeds actual decline in market value, appropriately deflated.[28] The effect of excess depreciation is to defer taxation of the excess until the asset is sold or to cancel tax altogether if the asset is held until the owner dies.[29]

By definition, consumption of and investment in owner occupied housing are done by the same household; these two activities are divided with respect to rental property. Consequently, whether any tax benefits on rental property accrue to the owner or to the renter is more significant than is the same question with respect to owner occupied property. If rental markets are competitive, tax savings from excess depreciation would be shifted from owners, who would receive the benefits in an accounting sense, to renters or to other owners of capital.

To the extent that tax depreciation formulas systematically provide owners with larger deductions than true economic depreciation would warrant, the analysis of the incidence of excess depreciation closely resembles that of tax benefits to homeowners. Assume that the cost of producing rental housing is constant (see Figure 2). Excess depreciation allowances will shift the supply of rental housing, causing the price of rental housing to decline, the amount of rental housing demanded to rise, and that of other housing to fall. If total housing demand rises, capital will have to be shifted from other industries to housing. This shift will occur through an increase in the yield of capital relative to payments of labor, a rise in the price of all housing relative to other commodities because of the capital intensity of housing, and an increase in factor earnings of wealth holders in general relative to recipients of wage and salary income.

Once again, disregarding the second and third effects, the magnitude of the direct impact on rents of excess depreciation depends on its size relative to rents and on

[27] Housing and all other structures were not eligible for the investment tax credit which was available on investments in most other tangible capital goods.

[28] Which price deflator to use when various prices are changing by different amounts is not altogether clear. For example, if prices of investment goods, in general, are rising faster than prices of all commodities, the size of capital gain (or loss) realized upon sale of an investment good, depends on whether the tax system is designed to tax changes in the investors command over investment goods or goods in general. On this problem, see W. Lewis.

[29] Before enactment of the Tax Reform Act of 1969, original owners of rental property could choose among various depreciation formulas, the most liberal consisting of double declining balance (*DDB*) and sum-of-the-year-digits (*SYD*). Original owners could substitute any other depreciation formula, even one which produced larger deductions in early years than *DDB* or *SYD*, if they could demonstrate that the alternate formula accurately reflected true depreciation. However, the purpose here is to explore the extent to which depreciation formulas lead to larger deductions than are justified by true economic depreciation. Consequently, only such formulas as *DDB* or *SYD* which were conclusively presumed to be acceptable are of interest here. Subsequent owners were limited to depreciation formulas producing deductions no more concentrated in early years than 150 percent declining balance, but they were freer to choose shorter estimated lives than original owners.

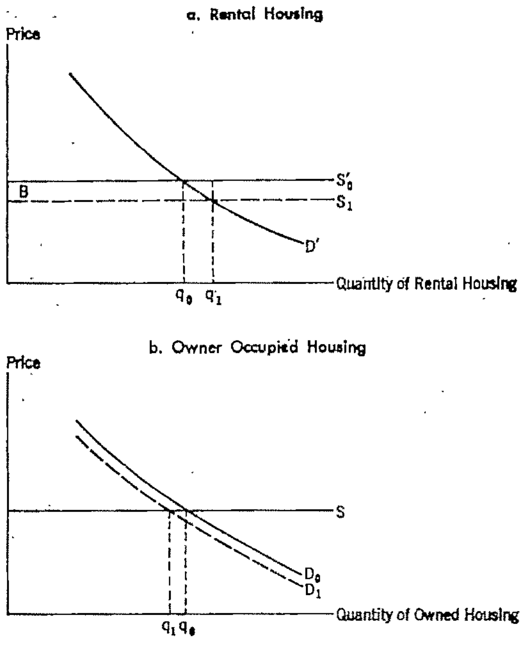


FIGURE 2. IMPACT OF EXCESS DEPRECIATION

TABLE 10—SUBSIDY EQUIVALENT OF TAX BENEFITS FROM EXCESS DEPRECIATION* ON RENTAL PROPERTY AS PERCENT OF GROSS RENT

True Depreciation	Marginal Tax Bracket			
	14	25	50	70
4.5%	0	0	0	0
4.0	0.4	0.8	1.6	2.2
3.0	1.3	2.3	4.7	6.6
2.0	2.2	3.9	7.8	10.9
1.0	3.1	5.5	10.9	15.3
0.0	3.9	7.0	14.1	19.7

* Property is assumed to have a useful life of 40 years. Gross rents are assumed to equal 16 percent of market price. Depreciation at double declining balance is taken to be the unweighted, undiscounted average of the first five years, 4.5 percent.

the marginal tax bracket of the owner. Unfortunately, virtually nothing is known about the rate of true economic depreciation on housing. Table 10 contains illustrative examples of the value of excess depreciation to taxpayers in various marginal brackets. The numbers in Table 10 are based on the presumption that there is *no* recovery of excess depreciation through income taxation at the time of a later sale.[30] If some portion of excess depreciation were recovered through taxation at the time of a later sale, the value of the subsidy would be smaller than that shown in Table 10. The diminution in value would be smaller the longer the deferral of sale and the higher the subjective discount rate of the taxpayer. In any case, however, the value of the subsidy on rental property through excess depreciation is almost certainly much smaller than that to homeowners, unless true economic depreciation is near zero and there is little or no recovery of excess depreciation at time of subsequent sale.[31]

III

The Internal Revenue Code contains massive tax subsidies for housing. The largest accrue to homeowners through exemption from taxation of net imputed rent and deductibility of mortgage interest and property taxes paid. Smaller benefits accrue to owners of rental housing to the extent that accelerated depreciation exceeds true depreciation on real estate by a greater margin than accelerated depreciation on other properties.

These tax subsidies affect resource allocation, income distribution, and the form of legal tenure in which housing is held. Consumption of housing services is con-

[30] Such complete avoidance would occur only if the owner died before selling or donated the property to charity. Much of the same result could occur if the owner traded the property for another, in which case, the "basis," or depreciated value of the original property would be carried over to the new one. This process could go on indefinitely. See M. David and R. Slitor.

[31] The tax benefits from excess depreciation to a high bracket owner of rental property, however, may easily be greater than would be the benefits to a low bracket household from owning the same property. The likelihood of this situation diminishes the higher the tax bracket of the occupant. Accordingly, tax factors help explain the fact that the fraction of households renting apartments or houses declines sharply with income.

siderably greater than it would be in the absence of tax benefits, particularly by homeowners. The magnitude of this effect is difficult to gauge, but is about 20 percent in the aggregate, and much greater for some groups.[32]

The impact on income distribution is large. In an accounting sense the benefits accrue primarily to upper income homeowners. These tax benefits are formally equivalent to direct housing subsidies combined with a tax system which treats income from housing as other income from capital is treated.

By increasing demand for housing services, a highly capital intensive commodity, tax benefits probably raise yields on capital in general. This increase affects all recipients of income from capital. Furthermore, it raises the price of capital intensive commodities such as housing, relative to the price of other commodities. Since tax benefits on rental property are so much smaller than those available to homeowners, the former may do little more than offset the price increasing effects, through capital costs, of the latter.

With respect to any conceivable policy objective, the pattern of tax benefits seems to be capricious and without rationale. Apart from the alleged, but unsubstantiated, benefits accruing to the community when households come to own their own homes, there appears to be no reason for subsidizing homeownership rather than other investments or the consumption of owned rather than rented housing services or of other commodities.[33] Even if one were to acknowledge that homeownership benefits society by making homeowners more stable or less antisocial than they otherwise would have been, the *pattern* of tax benefits is ill suited to that objective. Tax benefits provide largest benefits to recipients of larger than average income whose experience with wealth is typically not limited to their own houses. They provide negligible aid to low income households, most of whom have not been vouchsafed the salutory discipline of property management.

The obvious remedy for the problems raised in this paper is that homeowners be required to report gross imputed rent and expenses on their residence as other businesses must report gross income and expenses generated in producing other goods.

The implementation of such a procedure would raise some serious administrative problems. It would be necessary through appraisal to establish net imputed rent. At least two approaches could be used. First, direct appraisals of gross imputed rent could be prepared by appraisers, much as appraisals of fair market value are now made by property tax offices, and for banks, insurance companies, savings and loan associations and other financial intermediaries. From these estimates maintenance expense, depreciation, and interest would be deducted to reach net rent. Alternatively, the owner's equity interest could be estimated by deducting the balances of outstanding debt secured by the house from the appraised fair market value of the house. Net imputed income

[32] This estimate presumes a mean marginal tax rate among homeowners of 22 percent (in 1966 the mean marginal rate among all taxpayers was 19 percent), that imputed rent plus deductible expenses average five-eighths of gross rent and that the price elasticity of demand for housing is -1.5.

[33] A vast literature exists which purports to show the beneficent effects of good housing or of homeownership or the harmful effects of slums on households or on society. References to major contributions to this literature may be found in J. Rothenberg (p. 58) and A. Schorr. In this author's opinion, no study has shown both of the following: (a) that the beneficial effects of housing are due to housing itself rather than adequate income, i.e., that the composition, rather than the level, of consumption matters; (b) that correlations between homeownership and socially or personally desirable characteristics (or the absence of antisocial characteristics) are not the joint results of other psychological, sociological or economic characteristics. The issue of which way causation runs also is frequently troublesome.

could be estimated by multiplying owners equity by a rate of return equal to that on alternative investments the homeowner could have made, such as the interest rate on long-term government bonds, high grade corporate bonds, Regulation Q rates, or rates on some other form of security.

On balance, the latter method seems preferable since under the first procedure the homeowner, by discretionary behavior, could keep net rent artificially low. Each dollar spent for maintenance or noncapital improvements would reduce tax liability, thereby encouraging homeowners, particularly high bracket taxpayers, to make improvements in excess of those which they would choose to make apart from tax considerations. Moreover, since capital gains on owner occupied homes are treated even more favorably than are capital gains on other assets, the chance of even partial recovery through capital gains taxation at time of sale would be far from certain. This avenue of avoidance would be narrowed under the latter procedure. While the administrative problems are not trivial, they do not appear insuperable. The reliance which financial intermediaries place on property appraisals, suggests that appraisals of sufficient accuracy to support estimates of imputed rent can be made. The additional costs of making such appraisals would be substantial, however, unless appraisals for federal income tax and local property taxes could be joined. Such a course would generate major collateral benefits through improvement in the administration of the property tax.

Alternatively, deductions could be disallowed for mortgage interest, property taxes, or both. If deductions for mortgage interest were disallowed, homeowners would be encouraged to substitute borrowing on other assets for borrowing on their homes. To the extent that homeowners could substitute other credit instruments for home mortgages or had already paid off their mortgages, disallowance of deductions for mortgage interest would have no tax consequences. However the impact would be severe on markets for financial assets. Clearly, deductions for mortgage interest could not be disallowed in practice without major changes in laws regulating savings and loan associations.

None of the preceding problems would arise if deductions for property taxes paid were disallowed. No valuation problems would arise, nor would any tendencies to distort asset choice. Rather significant political consequences would ensue, however. Property tax increases would generate greater net costs for homeowners and aggravate the already sharp discontent over property taxation. Disallowance of property tax deductions would not appear to be politically acceptable unless it were combined with financial aid, through revenue sharing or other fiscal devices, for the municipal governments and school districts, thereby permitting large reductions in property tax rates. If complete reform in the taxation of homeowners is not adopted, then disallowance of deductions for property taxes should be included as part of any large scale program of revenue sharing, block grants, or other fiscal devices designed to strengthen revenues of local governments.

In summary, the program of complete reform seems the most attractive course, particularly if it were introduced as part of an overall tax reduction to offset the tax increases which homeowners would otherwise experience. The reform would develop a common interest between the federal government, municipal governments, and school districts in accurate property tax appraisal. It would reduce the substantial horizontal inequality resulting from the anomalous treatment of homeowners. And, at given tax rates, it would cause the personal income tax to generate substantially larger revenues which could

support direct expenditures for housing or other purposes, reductions in tax rates, or lower interest rates.

Appendix

The Brookings Tax File is a sample of 86,610 representative returns for the year 1966. These returns contain all information on deductions reported in the original tax return. Tax liability was computed initially exactly as computed by the filer, except that taxpayer errors were corrected. Next, deductions claimed by taxpayers for mortgage interest and property taxes were disallowed (Law 1). In most cases, taxable income rose by an equal amount and tax liability rose by the product of the change in taxable income and the relevant marginal tax rate (or rates). In some cases, disallowance of deductions for mortgage interest and property taxes caused total deductions to become smaller than the minimum standard or standard deductions, in which case the computation proceeded on the basis most favorable to the taxpayer.

The estimation of imputed rent proceeded in two steps. First, all tax returns on which deductions for mortgage interest or property taxes were made were assumed to have been filed by homeowners. On the basis of these returns, the total number of homeowner taxpayers *who itemized* is estimated by income class. Second, the number of *all* homeowner taxpayers in each income class was estimated by multiplying the total number of taxpayers in each income class by the fraction of homeowners in each income class as computed from Dick Netzer. The difference between the total number of homeowner taxpayers and the number of homeowner taxpayers who itemize is the number of homeowner taxpayers who claim standard or minimum standard deductions. This residual exceeded the number of taxpayers using the standard deduction in the top three income classes ($25,000–$50,000, $50,000–$100,000, and over); 85 percent, 90 percent, and 95 percent, respectively, of taxpayers claiming standard deductions in these classes were assumed to to be homeowners.

The amount of net imputed rent assigned to each household was computed as follows. From data in Projector and Weiss, the percent of households in each income class with equity interest in owned homes of varying amounts in 1963 was estimated. These estimates were increased by 18.76 percent, the amount by which median house value rose between 1963 and 1966 according to Katona (p. 62). Imputed net rent was set at 4 percent or 6 percent of estimated equity. Homeowner taxpayers in each income class were randomly assigned one of the resulting amounts of net imputed rent, with the probability of assignment of a specific amount dependent on the proportion of all households in that income class falling in that category. The actual distribution of homeowners by income and net rent class is available from the author on request.

The nine income intervals reported in text tables were dictated by the income intervals used in Projector and Weiss. Data on the distribution of benefits from deductibility of mortgage interest, property taxes, or both in combination have also been computed for a more detailed breakdown by thirty income classes.

REFERENCES

M. David, *Alternative Approaches to Capital Gains Taxation*, Washington 1968.

E. Denison, *Why Growth Rates Differ*, Washington 1967.

R. Goode, "Imputed Rent of Owner Occupied Dwellings Under the Income Tax," *J. Finance*, Dec. 1960, *15*, 504–30.

R. Hartman, "Demand for the Stock of Non-Farm Housing," unpublished doctoral dissertation, Harvard Univ. 1963.

H. Houthakker and L. Taylor, *Consumer Demand in the United States, 1929–1970—Analysis and Projections*, Cambridge 1966.

G. Katona et al., *1966 Survey of Consumer Finances*, monograph No. 44, Sur. Res. Ctr., Univ. Michigan 1967.

D. Laidler, "Income Tax Incentives for Owner Occupied Housing," in A. Harberger, ed., *Taxation of Income from Capital*, Washington 1969, pp. 50–76.

T. H. Lee, "The Stock Demand Elasticities of

Non-Farm Housing," *Rev. Econ. Statist.*, Feb. 1964, *46*, 82–89.

W. Lewis, "Depreciation and Obsolescence as Factors in Costing," in J. Meij, ed., *Depreciation and Replacement Policy*, Amsterdam 1961, pp. 15–45.

A. Manvel, "Trends in the Value of Real Estate and Land, 1956 to 1966," in *Three Land Research Studies*, National Commission on Urban Problems Res. Rept. No. 12, Washington 1968.

P. Mieszkowski, "On the Theory of Tax Incidence," *J. Polit. Econ.*, June 1967, *75*, 250–62.

———, "The Property Tax: An Excise Tax or a Profits Tax," mimeo.

R. Muth, "The Demand for Non-Farm Housing," in A. C. Harberger, ed., *The Demand for Durable Goods*, Chicago 1960, pp. 29–96.

———, *Cities and Housing*, Chicago 1969.

D. Netzer, *Economics of the Property Tax*, Washington 1966.

O. Oldman and H. Aaron, "Assessment-Sales Ratios under the Boston Property Tax," *Nat. Tax J.*, Mar. 1965, *18*, 36–49.

A. Ott and D. Ott, "Simulation of Revenue and Tax Structure Implications of Broadening the Federal Income Tax Base," in A. B. Willis, ed., *Studies in Substantive Tax Reform*, Chicago 1969.

D. Projector and G. Weiss, *Survey of Financial Characteristics of Consumers*, Board of Governors of the Federal Reserve System, 1966.

M. Reid, *Housing and Income*, Chicago 1962.

J. Rothenberg, *Economic Evaluation of Urban Renewal*, Washington 1967.

A. Schorr, *Slums and Social Insecurity*, U.S. Department of Health, Education, and Welfare, 1963.

J. Shelton, "The Cost of Renting Versus Owning a Home," *Land Econ.*, Feb. 1968, *44*, 59–72.

R. Slitor, *The Federal Income Tax in Relation to Housing*, Res. Rept. No. 5, The National Commission on Urban Problems, Washington 1968.

R. W. Tinney, "Taxing Imputed Rental Income on Owner-Occupied Homes," in A. B. Willis, ed., *Studies in Substantive Tax Reform*, Chicago 1969.

M. White and A. White, "Horizontal Inequality in the Federal Income Tax Treatment of Homeowners and Tenants," *Nat. Tax J.*, Sept. 1965, *18*, 225–39; also The Brookings Institution Reprint No. 114, with additional detail.

U.S. Internal Revenue Service, *Statistics of Income, Individual Income Tax Returns, 1966*, Washington 1968.

The Factor Price Frontier with Embodied Technical Progress

By David Levhari and Eytan Sheshinski*

It is known that any constant returns to scale production function in capital and labor has a dual, namely a relation between the wage rate and the rate of return on capital, called the Factor Price Frontier (*FPF*), (see Paul Samuelson and Michael Bruno). It is also known that in the presence of Harrod-neutral disembodied technical progress, the relation between the discounted wage rate, (the wage rate discounted with the efficiency factor) and the rate of return is invariant to changes of the rate of technical progress. Thus, if β is a constant rate of disembodied technical progress, W the nominal wage rate, $w=We^{-\beta t}$ the discounted wage rate and r the rate of return, the *FPF* is a relation $w=w(r)$, independent of β. The *FPF* gives w as a decreasing function of r, and has the property that its elasticity is equal to the ratio of the relative shares of capital and labor (see Peter Diamond).

We wish to analyze which of the above properties holds for the case where technical progress is at least partly embodied in capital equipment (see Robert Solow).

Using the neoclassical model with *ex ante* as well as *ex post* substitution between labor and capital, we observe that in steady-states:

1) given the rates of disembodied and embodied progress, w and r are negatively related as in the familiar case;

2) this relation is independent of the rate of disembodied technical progress;

3) it depends, in a negative way, on the rate of embodied technical progress!

4) the elasticity of w with respect to r is generally not equal to the ratio of relative shares.

I. *The Neoclassical Model with Neutral Embodied Technical Progress*

Let $L(\mathrm{t})$ be the number of laborers available at time t, and $K(\mathrm{t})$ be the quantity of capital equipment produced at time t. The rate of growth of labor is assumed to be constant

$$\text{(1)} \qquad \frac{dL(\mathrm{t})}{L(\mathrm{t})} = n$$

For simplicity, it is assumed that capital equipment is infinitely durable. The output, $Q(v,\mathrm{t})$, produced at time t by employing capital equipment of vintage v, is determined by $K(v)$ and the amount of labor allocated to this equipment, $L(v,\mathrm{t})$,

$$\text{(2)} \quad Q(v, \mathrm{t}) = F[K(v), \exp{(\alpha v + \beta \mathrm{t})} L(v, \mathrm{t})]$$

where F is the production function, α and β are the rates of embodied and disembodied technical progress, respectively. It is assumed that F exhibits constant returns to scale and diminishing marginal rate of substitution. Hence, we have that

$$\text{(3)} \qquad Q(v, \mathrm{t}) = g[\lambda(v, \mathrm{t})]K(v),$$

where

$$\text{(4)} \qquad \lambda(v, \mathrm{t}) = \frac{\{\exp(\alpha v + \beta \mathrm{t})\} L(v, \mathrm{t})}{K(v)}$$

is the labor-capital ratio in efficiency units, and $g(\lambda)$ is a shorthand notation for $F[1,\lambda]$.

We make the additional hypotheses that

* The Hebrew University of Jerusalem.

(5) $g(\lambda)$ is continuously twice differentiable;

(6) $g(\lambda)>0$, $g'(\lambda)>0$, $g''(\lambda)<0$, for all $\lambda>0$;

(7) $g(0)=0$, $g(\infty)=\infty$, $g'(\infty)=0$.

Finally, two alternative assumptions are examined:

(8) $g'(0)$ = (positive constant, say, =) 1,

$$g'(0) = \infty \tag{9}$$

Perfect competition assures that the available labor force at time t is allocated among the various vintages so as to equate the marginal product of labor to the prevailing wage rate, $W(t)$. Thus, at any t,

$$W(t) = \frac{\partial Q(v, t)}{\partial L(v, t)} \qquad \text{for } v \leqq t$$

Let $w(t)$ be the wage rate deflated by labor efficiency: $w(t) = W(t)(\exp -(\alpha+\beta)t)$ Then from the previous definitions we have that

$$w(t) \exp [\alpha(t - v)] = g'[\lambda(v, t)] \tag{10}$$

The wage rate $w(t)$ uniquely determines $\lambda(v,t)$, but we restrict the solution $\lambda(v,t)$ of (10) to be nonnegative. Consider first the case (9). In this case labor is allocated to *every* existing capital equipment, i.e. $\lambda(v,t)>0$ for each (v,t). We may therefore write

$$\lambda(v, t) = h[w(t) \exp \alpha(t - v)] \qquad v \leqq t \tag{11}$$

where $h = g'^{-1}$ (the inverse of g').

Let $k(t)$ be the ratio of capital equipment produced at time t to the amount of labor measured in efficiency units

$$k(t) = \frac{K(t)}{[\exp (\alpha + \beta)t]L(t)} \tag{12}$$

Then the condition of labor full-employment

$$L(t) = \int_{-\infty}^{t} L(v, t)dv \tag{13}$$

can be written, using (1), (3), (11), and (12)

$$1 = \int_{-\infty}^{t} h[w(t) \exp \alpha(t - v)] \cdot \exp[-(\beta + n)(t - v)]k(v)dv \tag{14}$$

Since h is a strictly decreasing function, ranging from infinity to zero, given the capital profile $k(v)$, the equilibrium wage rate is uniquely determined by (14).

It is assumed that a constant fraction $0<s\leq 1$ of total gross output, $Q(t)$, is saved. The amount of capital equipment produced at time t is then given by

$$K(t) = sQ(t) = s\int_{-\infty}^{t} Q(v, t)dv \tag{15}$$

or using intensive variables

$$k(t) = s\int_{-\infty}^{t} g(h[w(t) \exp \alpha(t-v)]) \cdot \exp\{-(\alpha+\beta+n)(t-v)\}k(v)dv \tag{16}$$

Equations (14) and (16) uniquely determine the path of capital accumulation for any given initial profile of the capital stock.

In a one-sector model we can measure capital goods in output units, and set the price of new capital equipment at unity. We thus have for every (v,t), $v\leqq t$,

$$p(v, t) = \int_{t}^{\infty} \rho(v, u)[\exp -\int_{t}^{u} r(z)dz]du, \tag{17}$$

where

$$\rho(v, u) = g(h[w(u) \exp \alpha(u - v)]) - w(u)[\exp \alpha(u - v)] \cdot h[w(u) \exp \alpha(u - v)] \tag{18}$$

is the rental earned at time u by a machine of vintage v, and where $r(z)$ is the rate of return at time z. Setting $p(t,t)=1$, equa-

tion (17) can be regarded as an integral equation for the unknown rate of return which itself is a function of time.

Now consider case (8). Under this assumption, only those vintages for which

$$g'(0) \geqq w(t) \exp \alpha(t - v)$$

are actually utilized. Then the active capital stock consists only of the newest machines. Let $m(t)$ be the age of the oldest equipment in use; namely

$$w(t) = \exp[-\alpha m(t)] \tag{19}$$

The active capital stock comprises the equipment constructed between $t-m(t)$ and t. The equilibrium conditions (14) and (16) now take the form

$$1 = \int_{t-m(t)}^{t} h[\exp \alpha(t-v-m(t))] \cdot \exp[-(\beta+n)(t-v)]k(v)dv \tag{20}$$

and

$$k(t) = s\int_{t-m(t)}^{t} g(h[\exp \alpha(t-v-m(t))] \cdot \exp[-(\alpha+\beta+n)(t-v)]k(v)dv \tag{21}$$

The equation to determine the rate of return has to be similarly rewritten to take into account the finiteness of the life of equipment.

It is possible to show that there is a particular path of capital accumulation for which the amount of new equipment per unit of labor measured in efficiency units $k(t)$ remains constant. Along such a path, it can be shown, the wage rate in efficiency units $w(t)$, the economic lifetime $m(t)$, and the rate of return $r(t)$ all remain constant.

Substitution of $k(t)=k$, $w(t)=w$, $r(t)=r$, and $x=t-v$, into (14), (16), and (17) yields

$$1 = k\int_0^\infty h[w \exp \alpha x][\exp -(\beta+n)x]dx \tag{22}$$

$$1 = s\int_0^\infty g(h[w \exp \alpha x]) \cdot [\exp -(\alpha+\beta+n)x]dx \tag{23}$$

$$1 = \int_0^\infty \{g(h[w \exp \alpha x]) - wh[w \exp \alpha x] \exp \alpha x\} \exp(-rx)dx \tag{24}$$

which are the equations that determine the steady state value of k (equation (22)), w (equation (23)) and r (equation (24)).

For the model with a finite life of capital in use, the corresponding steady state equations are

$$1 = k\int_0^m h[\exp \alpha(x - m)] \cdot [\exp(-(\beta + n)x)]dx \tag{25}$$

$$1 = s\int_0^m g(h[\exp \alpha(x - m)]) \cdot \exp[-(\alpha + \beta + n)x]dx \tag{26}$$

$$1 = \int_0^m \{g(h[\exp(\alpha(x - m))]) - h[\exp(\alpha(x - m))] \cdot \exp[\alpha(x - m)]\} \exp(-rx)dx \tag{27}$$

which are the equations that determine the steady-state values of k (equation 25)), m (equation (26)), and r (equation (27)). Given m, the wage rate is determined by equation (19) above: $w=\exp(-\alpha m)$

It should be pointed out that with embodied technical change, the rate of interest r may become negative. The possibility of a negative rate of return has been shown by Levhari-Sheshinski, and reflects the obsolescence caused by high savings or a high rate of embodied technical progress.

II. *The Factor Price Frontier* (*FPF*)

The steady-state solutions of the model, m, w, k, and r, are all functions of the parameters s, α, β, and n.

Given the technical progress rates α and β and the population growth, one can trace the values of w and r for various values of the savings rate s, with s as the parameter. The resulting curve is the *FPF*.

Let us now analyze the characteristics of the *FPF*. Consider first the model described by equations (22)–(24). Define the function

$$\phi(w, \alpha, r) = \int_0^\infty g(h[w \exp(\alpha x)]) - h[w \exp(\alpha x)]w \cdot \exp(\alpha x) \exp(-rx)dx \tag{28}$$

By (24), we have

$$\phi(w, \alpha, r) = 1 \tag{29}$$

Thus, (29) defines implicitly the *FPF*. The first observation to be made is that the *FPF* is independent of the rate of disembodied progress β.[1]

From (29), the slope of *FPF* can be calculated

$$\frac{dw}{dr} = -\frac{\phi_r}{\phi_w} \tag{30}$$

It can be shown that

$$\phi_r = -\int_0^\infty x[g(h) - hg'(h)] \exp(-rx)dx < 0$$

$$\phi_w = -\int_0^\infty h(\exp -(r-\alpha)x)dx < 0$$

Hence

$$\frac{dw}{dr} < 0$$

[1] Notice that this independence relates to the wage rate in efficiency units, w. The undiscounted wage rate $W(t) = w \exp(\alpha+\beta)t$ clearly depends on β. In fact, given α and r, an increase in β increases W proportionately.

It is well known that in the case of purely disembodied technical progress, the elasticity of *FPF* at any point is equal to the ratio of profits to the wage-bill (or ratio of relative shares). This property does not hold for the case of embodied technical progress.

From (22)–(24), it can be shown that the relative share of wages, denoted by $1-\pi$ is equal to

$$1 - \pi = \frac{w \int_0^\infty h[w \cdot \exp(\alpha x)] \exp[-(\beta + n)x]dx}{\int g(h[w \cdot \exp \alpha x]) \exp[-(\alpha + \beta + n)x]dx}$$

and, the ratio of relative shares is therefore given by (31). On the other hand, the elasticity of the *FPF* is given by (32). It seems that in general (32) is not equal to (31).[2] (See p. 811)

Let us now calculate the effect of a change in the rate of embodied progress on the *FPF*.

For a given rate of return r, we have

$$\frac{dw}{d\alpha} = -\frac{\phi_\alpha}{\phi_r}$$

Since

$$\phi_\alpha = -w \int_0^\infty xh(\exp -(r-\alpha)x)dx < 0$$

we find that

$$\frac{dw}{d\alpha} < 0 \tag{33}$$

Equation (33) means that the *FPF* moves *inwards* as the rate of embodied technical progress increases! See Figure 1.

One implication of this phenomenon is the following. Given the overall rate of

[2] Even in the Golden Rule case, $r = \alpha + \beta + n$, the numerator in (31) is not equal to that in (32).

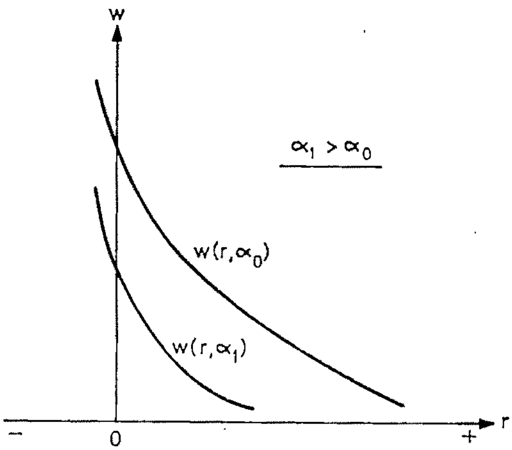


FIGURE 1

technical progress $\alpha+\beta$, the higher α (and smaller β), i.e. the higher the rate of embodied relative to disembodied technical progress, the lower will be the wage rate[3] at a given rate of return on investment.

This suggests another implication. From data on the overall growth rate $(\alpha+\beta+n)$ and the growth rate of population (n), one obtains a figure on the total rate of technical progress $(\alpha+\beta)$. From data on the nominal wage rate (W) it is then possible to calculate the discounted wage rate (w). Choosing a particular production function, data on w and r enables estimation of the FPF parameters including the rate of embodied progress α! Thus, contrary to some assertions, such as Jorgenson, in principle at least, it is possible to identify embodied from disembodied technical progress. This will come up clearly in the Cobb-Douglas example.

III. *The Cobb-Douglas Case*

Consider the Cobb-Douglas production function:

$$g(h) = h^{1-\gamma}, \tag{34}$$

where $0<\gamma<1$ is the elasticity (share) of capital. Using the definition of h and substituting in (24) one gets the FPF:[4]

$$w=(1-\gamma)\left[\alpha\left(\frac{1-\gamma}{\gamma^2}\right)+r\right]^{-\gamma/1-\gamma} \tag{35}$$

whose form is as in Figure 2. Here we have an example that the rate of interest r may become negative. Specifically, as r approaches $-\alpha(1-\gamma/\gamma)^2$, w approaches infinity.

As was asserted before, the elasticity of w with respect to r is seen not to be equal to $\gamma/1-\gamma$, the ratio of the relative shares of capital and labor. Also, given r, the wage rate w decreases with α. It is ob-

$$\frac{\pi}{1-\pi}=\frac{\int_0^\infty\{g(h[w\cdot\exp(\alpha x)])-g'(h[w\cdot\exp(\alpha x)])h[w\cdot\exp(\alpha x)]\}\exp[-(\alpha+\beta+n)x]dx}{\int_0^\infty h[w\cdot\exp(\alpha x)]g'(h[w\cdot\exp(\alpha x)])\exp[-(\alpha+\beta+n)x]dx} \tag{31}$$

$$\frac{-rdw}{wdr}=\frac{r\int_0^\infty x\{g(h[w\cdot\exp(\alpha x)]-g'(h[w\cdot\exp(\alpha x)])h[w\cdot\exp(\alpha x)]\}\exp(-rx)dx}{\int_0^\infty h[w\cdot\exp(\alpha x)]g'(h[w\cdot\exp(\alpha x)])\exp(-rx)dx} \tag{32}$$

[3] Since $\alpha+\beta$ is given, this holds for the discounted wage, w, as for the undiscounted $W=w\exp(\alpha+\beta)t$.

[4] See Levhari and Sheshinski.

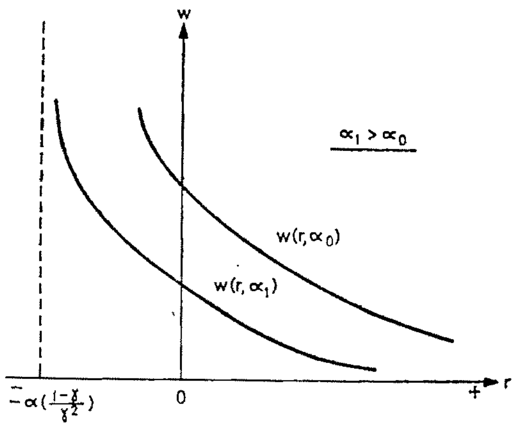


FIGURE 2

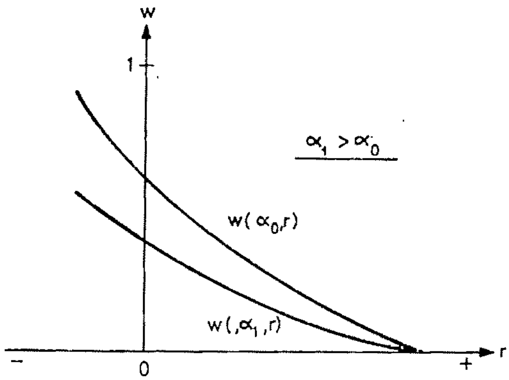


FIGURE 3

served that α and β are separately identified.

IV. *The Cases with Finite Lifetime of Capital*

All the above results apply to the model (25)–(27) in which the lifetime of capital is finite.

Using the relation $w=\exp(-\alpha m)$, the definition of ϕ changes here slightly as shown in (36) where, again by (27) $\phi(w, \alpha, r)=1$, which gives implicitly the *FPF*. The difference from the previous model is only that the upper limit of the integral in ϕ depends on w. But this leaves the values of the various derivatives (dw/dr given α, $dw/d\alpha$ given r, etc.) unchanged since the value of the integrand at the upper limit (the marginal vintage) is always zero.

As a further example, let us consider the fixed coefficients vintage model, brought forward in Solow et al. With a constant output-capital ratio, say one, equation (36) now has a simple form (see Diamond p. 368), and can be directly calculated.

$$(37) \qquad \phi(w, \alpha, r) = \int_0^{-1/\alpha \cdot \log w} [1 - w \cdot \exp(\alpha x)] \exp(-rx)dx = \frac{1}{r}[1 - w^{r/\alpha}] - \frac{w}{r-\alpha}\cdot[1 - w^{r-\alpha/\alpha}]$$

Putting $\phi(w,\alpha,r)=1$, (37) gives the *FPF* for this case. It is impossible, of course, to solve directly for w or r, but the general characteristics can be analyzed. The *FPF* is downward sloping and convex.[5]

For any α, the maximum rate of interest is always one. Given α, there is always a maximum wage rate w that corresponds to the highest savings rate $s=1$.[6] The rate of interest may well become negative at high values of w.

$$(36) \qquad \phi(w, \alpha, r) = \int_0^{-1/\alpha \log w} \{g(h[w\cdot\exp(\alpha x)]) - g'(h[w\cdot\exp(\alpha x)])h[w\cdot\exp(\alpha x)]\} \exp(-rx)dx$$

[5] In general, the sign of d^2w/dr^2 is indeterminate.

[6] The maximum w, $\bar{w}$, corresponding to $s=1$, is given by $\bar{w}=[1-(\alpha+\beta+n)]\exp 1/(\alpha+\beta+n)$. See Solow et al. (p. 92).

REFERENCES

M. Bruno, "Fundamental Duality Relations in the Pure Theory of Capital and Growth," *Rev. Econ. Stud.*, July 1969, *36*, 39–54.

P. A. Diamond, "Technical Change and the Measurement of Capital and Output," *Rev. Econ. Stud.*, Oct. 1965, *32*, 161–68.

D. W. Jorgenson, "The Embodiment Hypothesis," *J. Polit. Econ.*, Feb. 1966, *74*, 1–12.

D. Levhari, and E., Sheshinski, "The Relation Between the Rate of Return and the Rate of Technical Progress," *Rev. Econ. Stud.*, July 1969, *36*, 363–79.

P. A. Samuelson, "Parable and Realism in Capital Theory, The Surrogate Production Function," *Rev. Econ. Stud.*, June 1962, *29*, 193–206.

R. M. Solow, "Investment and Technical Progress," in K. J. Arrow et al., eds., *Mathematical Methods in the Social Sciences*, Stanford 1960, pp. 89–104.

———, J. Tobin, C. C. Weizsacker, and M. Yaari, "Neoclassical Growth with Fixed Factor Proportions," *Rev. Econ. Stud.*, Apr. 1966, *33*, 79–116.

Resource Allocation with Increasing Returns to Scale

By Trout Rader*

A major problem of economic theory is to obtain necessary and sufficient conditions for efficient allocation of resources and distribution of products. In the absence of increasing returns there is producer efficiency if, and only if, industries separately maximize profits, given product prices and factor wages (see T. C. Koopmans 1951). In the absence of increasing returns and with convex and continuous preferences, there is consumer efficiency (or Pareto optimality) if, and only if, there is a competitive equilibrium in both production and consumption (with respect to some wealth distribution), (see Kenneth Arrow 1951, Gerard Debreu, Koopmans and A. Bausch, and Rader). Under similar circumstances, there is consumer efficiency if, and only if, consumers make any advantageous bargains among pairs given the trades with others (Rader). None of these results can be extended to the case of increasing returns to scale in one or more industries. In view of the importance of economies of scale in public utilities and many oligopolistic industries, this is a major deficiency of welfare economies.

Many of the propositions stated below are known for the case of constant or decreasing returns to scale. The chief novelty is that they are easily proved for the case of increasing returns. Although some of the examples may be familiar to the reader they are included for the sake of completeness.

I. *Preliminaries*

The succeeding terminology and discussion takes the first essay of Koopmans (1957) as a point of departure.

With regard to producer efficiency in any given time period, there are fixed quantities of factors of production. A (net) output is ***producer efficient*** if there is no other (feasible) (net) output giving at least as much of all commodities and more of one. It is ***consumer efficient*** (Pareto optimal) if it can be distributed in such a manner that no consumer can be bettered (by another production and distribution) without hurting others. In comparison with a production distributed among consumers in a particular way, another production is ***consumer superior*** provided the output can be distributed so as to improve the position of at least one consumer without deteriorating that of any other.

The main results are the titles of the sections following. For the most part, they depend upon a simple analogy between production functions and utility functions. There is producer efficiency if, and only if, industries try to maximize production subject to their technology and with the option of trade of factors with other industries, with fixed factor prices. There is producer efficiency if, and only if, factors are fully employed and industries minimize factor costs for given output. There is producer efficiency if, and only if, all factors are employed and marginal rates of substitution are equalized.

* Department of economics, Washington University, St. Louis. The research for this paper was financed by the National Science Foundation, NSFGS 1981. I have benefited from correspondence with Lloyd Shapley, James Friedman, John Chipman, Herbert Scarf, Theodore Bergstrom, and the referees.

There is consumer efficiency only if price equals marginal costs. If the output possibility set is convex, there is consumer efficiency if, and only if, the product is distributed according to principles of competitive pricing and price equals marginal costs. Convexity of the output set is neither assured nor excluded by increasing returns to scale. There is consumer superiority of production over one distributed to competitive consumers only if the production leads to a greater *GNP*. If the output set is convex, there is consumer efficiency if, and only if, distribution is competitive and *GNP* is maximized. Under increasing returns and with more than one good, the core may not exist.

As with most propositions, the above are not true without qualifications. In this section, there is a list of the assumptions which appear through the following exposition. Also, there are some simple results useful at a later point.

It is assumed that there are r industries producing n goods from m factors. For concreteness, let factor quantities be denoted $y(t)$ in time t, $y(t)$ a vector with m components, and let (net) outputs be denoted $x(t)$, a vector with n components. In any given time period factors are fixed in quantity. Over time, they are accumulated and/or deteriorate at rates which cannot be controlled except through production from other factors. Some of the coordinates of $x(t)$ refer to increments in factors. Other commodities which are for consumption only may also accumulate and deteriorate. Their stocks can be denoted $z(t)$, a vector with $n-m$ components. Both factors and non-factors are called *commodity stocks* of which there are n, $(y(t), z(t))$. There is the law:

$$(y(t+1), z(t+1)) = (y(t), z(t)) + x(t)$$

One path of commodity stocks would be producer superior to another if in any time period it gave at least as much of all commodity stocks as the other and for some period, and for some commodity stock, it gave more. Formally, $\hat{c}=(\hat{y}(t), \hat{z}(t), t=0,1, \ldots)$ is producer superior to $c=(y(t), z(t), t=0,1, \ldots)$ if for all $i, \hat{c}_i \geq c_i$ but $\hat{c} \neq c$. It would then follow that whenever the two started from the same initial quantities of commodity stocks, there would be a first time period such that one was superior, i.e. a first time for which

$$(\hat{y}(t), \hat{z}(t)) \geq (y(t), z(t)),$$

but

$$(\hat{y}(t), \hat{z}(t)) \neq (y(t), z(t))$$

In the preceding time period, both had the same factors (and other commodity stocks) and since only production can overcome the natural forces of deterioration, the one must have produced more. The other utilizes its inputs inefficiently at that time. Formally,

$$(\hat{y}(t), \hat{z}(t)) = (y(t-1), z(t-1)) + \hat{x}(t-1)$$
$$\geq (y(t), z(t)) = (y(t-1), z(t-1)) + x(t-1),$$

if, and only if,

$$(\hat{y}(t), \hat{z}(t)) - (y(t), z(t))$$
$$= \hat{x}(t) - x(t) \geq 0$$

Remark 1. Producer efficiency over time results if, and only if, there is producer efficiency in each time period separately.

All results following will refer only to the given time period, although they are related to the case of factors variable over time. The time notation is no longer required and use is made only of the variables, factor inputs, y, and net outputs, x.

Production is *interindustry* if the following conditions hold.

(i) Commodities are produced in r industries with output vectors x^i (negative components indicate raw material inputs; positive components, outputs). By convention, goods and industries are related so that $x_i^i > 0$. In many applications

the i^{th} good is not produced in positive quantities by any other industry, whereupon $r \leq n$. Production is by functions $f^i(y^i)$, where y^i is the vector of the quantities of factors used in the i^{th} industry and $f^i(y^i)$ is the amount of the i^{th} good produced, net in the industry but gross in the economy. It is assumed that f^i is continuous, i.e. small decreases in inputs lead to small decreases in outputs.

(ii) For the community, the *net output*, x, is obtained by summing industry outputs, x^i, which are vectors including the negative quantities representing the use of intermediate products. If more than two goods are produced in positive quantities in a given industry, they are joint outputs. The summation over industries of only positive components of x^i gives *gross output*. It will be assumed that no increase in all net outputs can be attained simultaneously with decreasing gross outputs in some industry. To assure this, one cannot increase the quantity of j produced in industry i and also reduce the raw materials requirements or the output of joint products, nor can one produce a good in positive quantities in two industries. Therefore, $|x_j^i|$ is a non-decreasing function of x_i^i. Also, there are implicit restrictions on the size of the negative x_j^i relative to x_i, more or less equivalent to assuming the productivity of a Leontief matrix.

(iii) Hereafter, the word factor of production is applied only to producer goods, namely those which can be utilized to produce some good in some industry. It is assumed that small increases in factors lead to small increases in output.

Summarizing,

$$x_i^i = f^i(y^i), \tag{1}$$

$$x^i = x^i(x_i^i), \tag{2}$$

$$x = \Sigma x^i, \tag{3}$$

$$y = \Sigma y^i \tag{4}$$

An example of (2) is the Leontief system where,

$$x^i = a^i x_i^i$$

a^i a vector of n constants and the square matrix relating the first r goods is a productive Leontief matrix. Other applications are not excluded, but this one can be regarded as basic.

Production is *quasi-concave* provided the production functions are quasi-concave:

$$f^i(\mathrm{t}y^i + (1-\mathrm{t})\bar{y}^i) \geq \min\,(f^i(y_i), f^i(\bar{y}^i)$$

whenever $0 \leq \mathrm{t} \leq 1$. The meaning of quasi-concavity is that the iso-product curves (or hyper surfaces) are convex to the origin. In effect, as the quantity of j increases, the marginal rate of substitution of i for j, when defined, is not increasing in absolute value (not decreasing in actual value). This does not exclude increasing, constant, or decreasing returns to scale since the returns to scale are determined by the distance between the iso-product curves (hypersurfaces).

Factors are productive in an industry whenever their increase increases output in that industry. Production is *factor connected* if for any industries i and k, there are industries q_m, $m=1, \ldots, p$, and associated factors $j(q_m)$ such that

$$q_1 = i,$$
$$q_p = k,$$

and $j(q_m)$, $j(q_{m-1})$ are utilized in positive quantities and are productive in the $q_m{}^{\text{th}}$ industry. In particular, every industry produces a positive (gross) output.

If labor is an indispensable factor in each industry and every industry is in operation, there is factor connectedness. On the other hand, if by changing the order of industries or factors, one has a matrix

$$(y_j^i) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

then production is not factor connected. Industries can be divided into two groups who use no factors in common. For example, there might be two industries, steel and coal, the one using only capital and the other only labor. Suppose there is idle capacity in steel because of an insufficient supply of coal, which goes into steel as an intermediate product. Then output of steel could be increased but only if less coal were available for other uses. A situation with idle steel capacity is efficient because more steel requires less coal, net. This state of affairs would make a theory of efficiency quite cumbersome.

LEMMA 1. *If production is interindustry and factor connected, then whenever one can increase gross output in one industry without decreasing it in others, we can increase gross output in all industries.*

PROOF:

For some industry, i, reduce gross output x_i^i, and utilize the factors released in another, j. Then transfer appropriate resources out of j into another, k, without reducing the output of the one to its former level, etc. By factor connectedness, starting from any industry, we can find a chain leading to an increased output in any other industries. In this way, the output of any (and hence all) industries can be increased by small quantities.

LEMMA 2. *If production is interindustry and factor connected, a net output is efficient if, and only if, the corresponding gross output is efficient.* The proof of the lemma would break down if more net outputs did not require more intermediate products.

PROOF:

Sufficiency can be seen by considering an increase in one commodity in its gross output. By Lemma 1, we can increase the gross output of all commodities. By producing only small amounts of one good and also enough of other goods to meet the one's intermediate product requirements, possibly leaving factors idle, there results a net output with more of the one and no less of the others. Repeated increases in net output from added factors will allow one to employ all factors, as required by (4). Here continuity of f^i must be applied to an increasing sequence of factors employed to show that in the limit the net output is still larger than in transition.

Actually, there may be a maximal nest of factors employed, increasing in quantity and whose limit will exhaust the factors available. The reader can consult any text book on General Topology for details on maximal nests.

For necessity, the above argument implies that a net output which can be increased in one component without decreasing others can also be increased in all components. By hypothesis this is possible only if gross outputs increase.

COROLLARY. *If production is interindustry and factor connected, the possibility of an increase in net output in one commodity with no decrease in others implies the possibility of an increase in the net output of all commodities.*

The proof is immediate from Lemmata 1 and 2.

LEMMA 3. *If production is interindustry and factor connected, a net output is efficient only if all factors are employed.* The lemma might fail if production were not factor-connected.

PROOF:

As in the proof of sufficiency in Lemma 2, extra factors can be used to increase the gross output of one industry, freeing factors for use elsewhere. There is a collection of such transfers which will increase all gross outputs, and by possibly not employing some factors one can actually increase a single output. Indeed one can continue this process, using up all the factors and increasing outputs, just as in the proof of Lemma 2. Therefore, the original production was not efficient.

Remark 2. If production is interindustry

Evidently, differential wages between industries in some but not all factors leads to inefficiency.

PROOF:

The necessity is immediate from Theorem 2 and from the observation that cost minimization with respect to w leads to equality between the marginal rates of substitution of j for k and w_j/w_k, whenever both factors are used. Otherwise, if j is not used, the marginal rate of substitution must not exceed w_j/w_k.

Sufficiency follows from the fact that a wage vector can be formed from the equality of marginal rates of substitution, by letting $w_i=1$ and then deducing the other wages as equal to the marginal rate of substitution of factors used in conjunctions with i, etc. They will be all nonnegative and for factors actually used, the wages will be positive. Hence, the wages will define a hyperplane equal to the set of $\bar{y}^i$ for which $y^i w=\bar{y}w$, which is tangent to F_i at y^i. Since F_i is convex, the hyperplane must bound F^i, and the industry minimizes costs with respect to w.

V. *Price Equals Marginal Cost and Competitive Consumer Choice Are Necessary But Not Sufficient for Consumer Efficiency*

As indicated in the preceding sections, monopoly may be compatable with producer efficiency, but it will be seen that monopoly is incompatable with consumer efficiency (see Rader). Consumer efficiency requires that commodities be produced in desired quantities as well as be produced efficiently.

It is well known that if consumers consume commodity bundles which are efficiently distributed, it is as if each consumer were optimizing preferences subject to a budget constraint. If the set of feasible outputs is convex, then the production sector should maximize the value of net output, $\sum p_i x_i$, which is *GNP*. This is

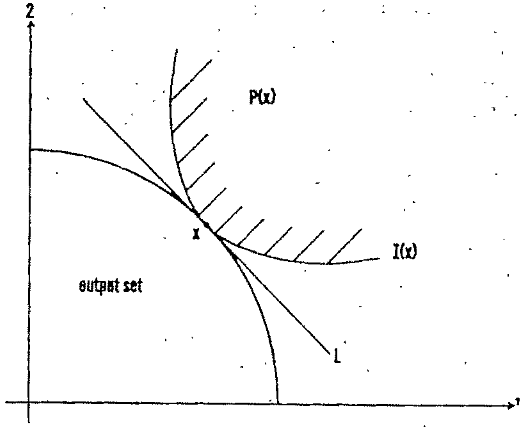


FIGURE 1

illustrated in Figure 1 for two commodities and one consumer who has indifference curve, $I(x)$, set of consumptions preferred to x, $P(x)$, and budget line L. Hence, output is consumer efficient if, and only if, consumer choice is competitive and *GNP* is at a maximum.

In many cases, under increasing returns in several of the industries, the production set is not convex, as illustrated in Figure 2, where industry 2 displays increasing returns. The consumer still chooses according to a budget constraint, but *GNP* is no longer maximized. Nevertheless, the line L is tangent to the output set. The line L' is also tangent to the output set and has the same relative prices as L. Evidently,

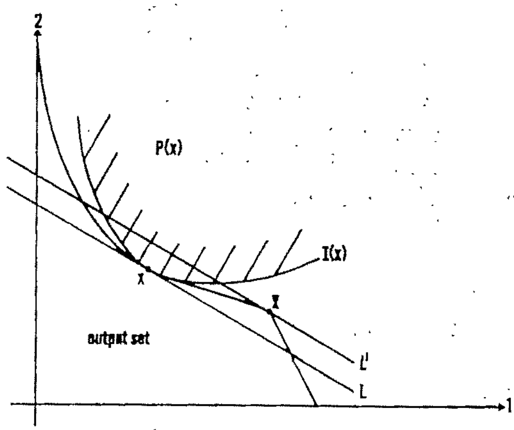


FIGURE 2

$\bar{x}$ is not consumer efficient even though, as will be seen, price equals marginal cost. Price equals marginal cost and competitive consumer choice are not sufficient for consumer efficiency.

Returning to the case of a convex output set, simply observe that under increasing returns there will be convexity if the loss of scale from the less intense operation of increasing returns industries is more than offset by economies of rearranging the proportions in which resources are used. In contrast, if there is but one factor of production, increasing returns is incompatable with the convexity of the output set.

Let $P_i(x^i)$ be the set of consumptions preferred by i to x^i. As is well known, x is consumer efficient if, and only if, the output set is disjoint from the set $\sum P_i(x^i)$. To establish the rule, price equals marginal cost, it is necessary to have uniqueness of the budget hypersurface bounding all $P_i(x^i)$ at x^i. This follows whenever for every y not in the bounding hyperplane, there is a t $\neq 0$ (possibly negative) such that $\sum x_i + ty$ is in $\sum P_i(x^i)$. In words, in the direction of production with greater *GNP*, there is a small change which can be distributed so as to improve the position of all consumers. In this case, it is said that $\sum P_i(x^i)$ is *directional dense.*

Production is *continuously differentiable* if all x^i are continuously differentiable in x_i^i. In what follows

$$\frac{dx^i}{dx_i^i} = \left(\frac{dx_j^i}{dx_i^i}\right)$$

is the rate of change of intermediate product usage with gross output, where $dx^i/dx^i = 1$. The derivative dy^i/dx^i is the rate of change in the least cost factor utilization with a change in gross output, and dwy^i/dx_i^i is the rate of change of factor costs with a change in gross outputs.

THEOREM 4. *If* $\sum P_i(x^i)$ *is directional dense, if the competitive consumer price system is* p, *and if production is interindustry, factor connected, quasi-concave, and continuously differentiable, then for a consumer efficient production with all products produced, there is a strictly positive wages system,* w, *such that*

(i) *industries minimize costs subject to the constant output constraint,*

(ii) $p\,(dx^j/dx_j^j)\ (\partial f^j/\partial_k^j) \leq w_k$ *with equality whenever* y^k *is used in positive quantities in industry* i *and*

(iii) $p_i = -\sum_{j\neq i} p_j dx_j^i/dx_i^i + dwy^i/dx_i^i$.

PROOF:

The wage system of Theorem 2 is used. Let $\partial x_i^i/\partial y_k^i$ be the marginal product of y_k^i so that $p(dx^i/dx_i^i)\ (\partial x_i^i/\partial y_k^i) = p\partial x^i/\partial y_k^i = w_k$, whenever $y_k^i > 0$. It is easy to see that this multiplies w by a positive constant. It is also easy to see that in the limit, small changes in output must have values which tend to non-positive numbers. For industries i and j with continuously differentiable f^i, f^j, for $y_k^i > 0$, $y_k^i > 0$, one can either transfer factor k from i to j or from j to i, so that the curve of change in net outputs must be tangent to the separating hyperplane. Hence, $p\partial x^i/\partial y_k^i - p\partial x^j/\partial y_k^j = 0$ or $p\partial x^j/\partial y_k{}^j = w_k$. For $y_k{}^i \neq 0$ and $y_k{}^j = 0$,

$$w_k = p\partial x^i/\partial y_k^i \geq p\partial x^j/\partial y_k{}^j$$

Also,

$$\begin{aligned} pdx^i/dx_i^i &= \Sigma_k(p\partial x^i/\partial y_k)(dy_k/dx_i^i) \\ &= \Sigma_k w_k dy_k^i/dx_i^i \end{aligned}$$

gives value-added equals marginal cost. Transfer raw materials costs to the other side of the equation, which gives the price equals marginal cost rule, (iii).

Remark 3. In Theorem 4,

$$p\partial x^i/\partial y_k \leq aw_k$$

and

$$p_i + \sum_{j \neq i} p_j dx_j^i/dx_i^i = adwy^i/dx_i^i$$

can be substituted in (ii) and (iii), $a>0$. Hence, factors of production need not be paid their marginal value products. It is necessary only that industries act as if they were paying the factors their marginal products.

PROOF:

If

$$p_i - \sum_{j \neq i} p dx_j^i/dx_i^i = adwy^i/dx_i^i,$$

one can simply increase wages to aw and apply Theorem 4.

Remark 3 established the conjecture of Lerner (ch. 9). It suggests the possibility of attaining optimality by equalizing the "degree" of monopoly over the economy by assuring everywhere the equality of average value-added with marginal factor costs. Simply tax competitive industries and subsidize monopolistic ones. Adjust the rates so that

(1) value-added bears the same proportion to marginal costs in all industries. The proportion may be any positive number, chosen so that

(2) total government revenues equal total government subsidies. If desired, another tax-subsidy scheme on factor incomes can be used to attain the distribution of income under pure competition. Incidentally, there are good welfare reasons for desiring the competitive distribution of benefits (Rader, Appendix B).

One case considered by Marshall is that where an industry has non-constant returns to scale but they are external to the firm. The firms can be pure competitors, and price equals industry marginal cost can be attained by an appropriate tax-subsidy scheme. The case of Cobb-Douglas production functions has been treated by John Chipman. In general, let there be n identical firms with marginal factor costs equal to average industry factor costs, AC. Let industry marginal factor costs be MC. Then the firms set average value-added (1+subsidy rate)$=AC$ which gives, average value-added$=MC$, if

$$\frac{AC}{1 - \text{subsidy rate}} = MC,$$

or

$$\text{subsidy rate} = \frac{AC}{MC} - 1$$

If the production function is homogeneous of degree n, $n\ MC=AC$ or subsidy rate $=n-1$, which is the classical result.

VI. *Larger GNP is Necessary but Not Sufficient for Consumer Superiority*

The title is almost trivial to prove. A consumption of greater or equal preference is at least as valuable as the going one. One actually better has higher value. Hence, the sum of consumption values is greater than current *GNP*. The reasons for insufficiency have been considered in the previous section.

VII. *Distributive Justice Under Increasing Returns*

The *core* is a production and allocation of consumption such that no person or group can do better for all its members relying solely upon its own resources. For an allocation to be in the core would seem to be an elementary requirement for distributive justice and it is called *Edgeworth optimal.* For the case of convex preferences and production, the core exists and, for large numbers of consumers, is equal to the competitive equilibrium (Debreu and Scarf).

Criteria have been defined for the existence of the core. The most general appears in Scarf (1967a) who gives an algorithm for computing it. Here, a much

simpler case appears where the payoff is the quantity of a single good produced in some industry from factors denoted x and y. For a given coalition of individuals, S, attainable output is limited by $f(\sum_{i \in s} x^i)$. The distribution of output between individuals is a *game with transferable utility*, since utility may be taken to be the quantity of the single commodity required, whereupon one can transfer utility from one individual to another.

The analysis of Lloyd Shapley applies:

THEOREM 6. *If* $x \geq y$, $\delta \geq 0$, *implies*

$$f(x+\delta) - f(x) \geq f(y+\delta) - f(y)$$

and $f(0)=0$, *then the game with transferable utility has a non-empty core.*

PROOF:

Allocate $u^1 = f(x^1)$ to 1, $u^i = f(x^i) - \sum_{j<i} u^j$ to i. No group can award more to all its members.

A function is *convex* if its negative is concave.

COROLLARY. If f is convex, $f(0)=0$ and there is only one factor of production, then the game with transferable utility has a non-empty core. This corollary, due to Shapley, has been proved with other methods by Scarf (1967b).

PROOF:

For $x \geq y$,

$$f(x+\delta) - f(x) = \int_x^{x+\delta} f' ds$$

$$\geq \int_y^{y+\delta} f' dy = f(y+\delta) - f(y)$$

since f' is non-decreasing in x, almost everywhere. Note that increasing returns to scale do not suffice for the existence of the core. In Figure 3a, there are increasing returns, but no convexity, which is illustrated in Figure 3b. Nevertheless, for 3a and for any factor endowment the core exists. A proof due to Shapley is based upon the fact that average output is increasing. Give to i the average output times the endowment of i. No subcoalition can block since it would have to be able to attain a higher average output. In view of this, it is probable that Theorem 6 can be improved.

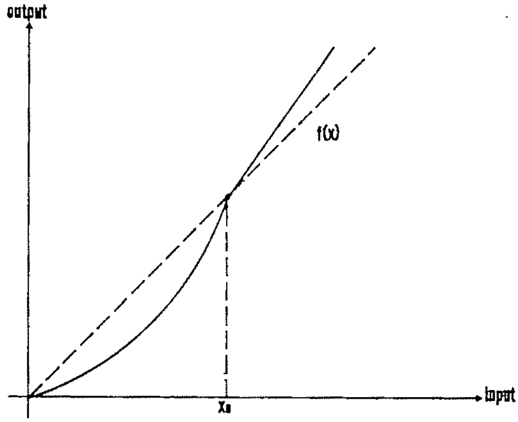


FIGURE 3a

Convexity appears to be the traditional concept of a public utility where marginal costs are decreasing and then constant. Theorem 5 and the corollary state that a public utility can assure enough to its suscribers to improve the position of any member or group relative to what it could get alone, provided it devotes the same resources to the public utility.

The caveat of the preceding paragraph is not a trivial one and it can be seen that

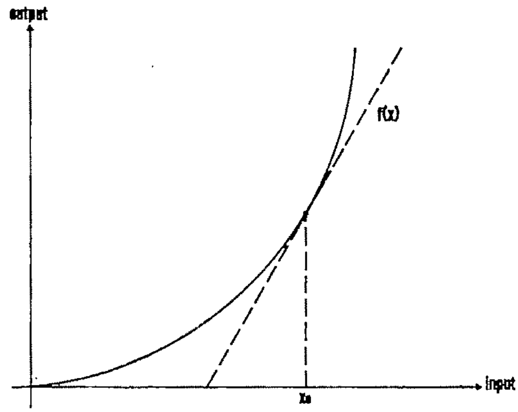


FIGURE 3b

there is no core in cases where several public utilities are competing for resources. To see why convexity does not assure the existence of Edgeworth optimal allocations, consider a three-person, three-good economy. Let consumer 1 like goods 2 and 3 and be indifferent between the proportions consumed of each. Similarly, let consumer 2 like 3 and 1 and consumer 3 like 1 and 2. Let the commodities be produced from a single factor by identical production functions, f, with increasing returns to scale, and let each consumer have an initial endowment of the factor, x. Then coalition $\{i, j\}$ will produce only commodity k, $i \neq j$, $k \neq i$, $k \neq j$. For the case of equal division of production, the coalition $\{1, 2, 3\}$ must assure each party at least $(1/2)f(2x)$ of the desired good. Obviously, producing all three goods is inefficient. Should we produce as much as $f(2x)$ of good i, then we must produce $(1/2)f(2x)$ of j, since otherwise the consumer not liking j can entice one of the others to join with him. Of course it is impossible to produce $f(2x)$ of i and $(1/2)f(2x)$ of j. Reducing the scale of output in good i will reduce the ratio of output to input in good i without bringing the input-output ratio in good j to an acceptable level. Of course, we cannot increase the scale of output of j without violating one person's benefit required to keep him from bidding away another from the arrangement. Formally, for increasing returns to scale, one cannot stimultaneously solve,

$$f(tx) + f((1-t)x) \geq 3/2f(2x)$$
$$f(tx) \geq 1/2f(2x),$$

and

$$f((1-t)x) \geq 1/2f(2x)$$

To remedy this one might imagine a production of one of the goods at the level $f(3x)$, the one to be so produced to be determined by a random device. However, this would be unacceptable if there is (sufficient) risk aversion on the part of the consumers.

A possible way out is for each person to regard coalitions as being assigned probabilities of formation. In this way, they apply the rule, should i make an offer to j to leave the coalition $\{1, 2, 3\}$ it is quite likely that k will do so also. Then, the payoffs for coalition (1, 2), (2, 3), and (3, 1) would be greatly reduced in threat against (1, 2, 3). With only slight risk aversion, a possible solution might be one third probabilities of each of three alternatives which were relatively unfavorable to each of the three parties. With strong risk aversion an alternative might be to give equal quantities of goods to all three parties. Whichever solution was accepted, it could always be blocked by some coalition of two. Without application of the categorical imperative, the resolution of conflict would seem impossible.

REFERENCES

K. J. Arrow, "An Extension of the Basic Theorems of Classical Welfare Economics," in J. Neyman, ed., *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley 1951, pp. 507–32.

——— and G. Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica,* July 1954, *22,* 265–90.

J. Chipman, "External Economies of Scale and Competitive Equilibrium," *Quart. J. Econ.,* Aug. 1970, *84,* 347–85.

G. Debreu, "Valuation Equilibrium and Pareto Optimum," *Proc. National Academy of Sciences of the U.S.A.,* 1954, *40,* pp. 588–592.

——— and H. Scarf, "A Limit Theorem on the Core of an Economy," *Int. Econ. Rev.,* Sept. 1963, *4,* 235–46.

T. C. Koopmans, "Analysis of Production as an Efficient Allocation of Resources," in Koopmans, ed., *Activity Analysis of Production and Allocation,* New York 1951, pp. 33–97.

———, *Three Essays on the State of Economic Science,* New York 1957.

——— and A. Bausch, "Selected Topics in Economics Involving Mathematical Reasoning," *SIAM Rev.*, I, 1959.

A. Lerner, *The Economics of Control*, New York 1946.

T. Rader, "Pairwise Optimality and Non-Competitive Behavior," in J. Quirk and A. Zarley, eds., *Papers in Quantitative Economics*, Vol. I, Lawrence, Kansas 1968, 101–27.

H. Scarf, (1967a) "The Core of an N-Person Game," *Econometrica*, Jan. 1967, *35*, 50–69.

———, (1967b) "Notes on the Cores of a Productive Economy," mimeo. 1967.

L. Shapley, "Notes on N-Person Games VII: Cores of Convex Games," RAND Corporation memorandum RM-4571-PR, Santa Monica 1965.

Dynamic Optimization and Economic Policy

By Eisuke Sakakibara*

Modern theories of economic growth have dealt mainly with the construction of deterministic mathematical models and the derivation of analytical results.[1] Rarely have growth theorists addressed themselves to questions concerning economic policy. This trend can be attributed to three factors. First, obtaining numerical solutions to complicated dynamic models has always been extremely tedious. Second, when the essential relations of a typical growth model have to be estimated statistically, problems for which no really satisfactory solutions exist are encountered. These include intercorrelation among the arguments of relations such as the production function, as well as deficiencies in the quality of the data itself (e.g., the stock of capital). Lastly, economists have tried to refrain from making specific assumptions about society's preferences.

Recent developments in modern control theory, especially in the area of computational methods, have alleviated the first problem. Extensive use of the digital computer now makes it possible to solve fairly complicated dynamic optimization problems numerically. Although the estimation and data problems remain, it would still be useful to apply the dynamic optimization techniques to the evaluation of some actual economic policies.

In order to derive meaningful results, it is necessary to extend the typical growth model by incorporating an aggregate demand sector and a foreign sector. In this way stabilization and balance of payments considerations can be taken into account simultaneously with questions concerning growth. The first section of the paper describes the formal model. Although our results are tentative, they can serve as good benchmarks to policy markers. Models of the kind presented here seem to have a promising future in the formulation and evaluation of economic policies.

I

This section presents a basic model which integrates the aggregate demand and foreign sectors with a conventional single-commodity growth model. For the sake of simplicity, we assume that policy makers monitor the economy and make their decisions on the basis of continual observations on two key variables, the investment-*GNP* ratio, s, defined as the ratio of gross investment to actual *GNP*, and the capacity utilization rate, cu, defined as the ratio of actual to full employment *GNP*. Although the authorities in decentralized economies such as Japan and the United States cannot effect changes in s and cu directly, they do have at their disposal a host of tools including the depreciation rate, investment tax credits, and the usual fiscal and monetary instruments by which they can achieve the target values of the investment-*GNP* and capacity utilization rates. It will be an interesting extension of this study if we can expand the model to include the more usual policy instruments.

* The author is currently with the Ministry of Finance, Japan. This paper is a part of his Ph.D. dissertation submitted to the University of Michigan in June 1969.

[1] See, for example, Kenneth Arrow and K. Shell. Koichi Mera has recently applied a less rigorous optimization model to the Japanese and *U.S.* data.

It is assumed that social preferences can be represented by an additive function of the discounted value of future consumption, the rate of unemployment, the rate of price inflation, and the deviation of the actual level of foreign reserves from the desired level.

The dynamic nature of the economy is captured in three "system equations"; one representing the accumulation of capital, one representing the formulation of inflationary pressure, and the other expressing the accumulation of foreign reserves.

Mathematically, we have the following dynamic optimization problem to solve:

(1) Maximize:[2]

$$W = \int_0^\infty e^{-rt}\{U[(1-s)(1-ucu)F(K,L;t)] + L(p) + Q(ucu) + G(FR^*)\}dt$$

(2) Subject to:

$$DK = (1 - ucu)sF(K,L;t) - \delta K$$

(3) $$Dp = g(ucu) + h(p)$$

(4) $$DFR^* = X(t) - M(t) + j(t)$$

where D is operator d/dt, W is the social welfare function, K is the capital stock, L the labor force, p the percentage change in the price index, and FR^* the difference between the actual level of foreign reserves FR and the desired level $\widetilde{FR}$. The variable ucu is the rate of excess capacity, defined as $1-cu$; s is the investment-GNP ratio, $X(t)$ the amount of exports, and $M(t)$ the amount of imports. The variable $j(t)$ is a time varying coefficient representing the movements of international capital transactions and the movements in the desired level of foreign reserves FR, r and δ are constants representing the pure time preference rate of society and the depreciation rate of capital, respectively. The functions U, L, Q, and G specify the utility or disutility associated with the levels of C, p, ucu, and FR^*, respectively. The partial derivatives satisfy the following conditions:

(5) $$U'(C) > 0,\quad U''(C) < 0,\quad L'(p) < 0,\quad L''(p) < 0,\quad Q'(ucu) < 0,\quad Q''(ucu) < 0,\quad G'(FR^*) < 0,\quad G''(FR^*) < 0$$

Equation (2) is a familiar one expressing the accumulation of capital where the function F is a production function. Equation (3) can be viewed as a reduced form equation for the wage-price sector of the economy.[3] Equation (4) defines the change in FR^*, which is the sum of the trade balance $(X(t)-M(t))$ and the combined effects of international capital transactions and changes in $\widetilde{FR}$, $(j(t))$.

The problem we have to solve is a typical control problem, where K, p, and FR^* are the "state variables," and s and ucu are the "control variables." Although interesting analytical results can be obtained for specific forms of the functions,[4] in this paper we shall concentrate on getting numerical results, analyzing the actual policies followed by Japan and the United States over the period 1952–67.

Although it is possible to solve the above problem using one of several algorithms which have been developed recently, we have instead chosen a two-step procedure which yields an approximation to the optimum. The two-step procedure has two advantages over the direct simultaneous solution.

First, given reasonable specifications of the functions of the model, it is possible to

[2] Note that the argument of the utility function U is not exactly the same as consumption. $(1-s)(1-ucu)F$ in the open economy is $C+X-M$. Thus the term $(1-s)(1-ucu)F$ already incorporates the trade surplus in it. The function G can be thought to take care of the additional utility the economy attaches to the trade surplus (or foreign reserves).

[3] For example, (3) can be considered to be a Phillips-type relation with a lag in prices and can be derived from two equations $Dw=w(ucu, p)$ and $Dp=(Dw, p)$, where Dw is the change in the wage rate.

[4] See my unpublished doctoral thesis.

derive a solution in feedback form for the second step. Since the direct method can only provide an open loop solution, and since our system equations neglect stochastic disturbances, direct solutions are not meaningful in a system where stochastic disturbances are dominant. On the other hand, the solution in feedback form can cope with this problem since the optimal solution is expressed as a function of the current values of the state variables which incorporate the stochastic disturbances of the past. Second, the actual process of decision making seems to take place in feedback form and in evaluating the economic policies of the two countries, we want to take this fact into consideration. In particular, we want to evaluate the performance of the actual stabilization policies independently of the growth policies, as coordination of these two is not well established in both Japan and the United States. It would not be fair to judge the stabilization policies on the assumption that they have been coordinated with the growth policies when in fact they have not. Elsewhere the author has shown that for a growth model incorporating a demand sector but no foreign sector, the equilibrium investment-*GNP* ratio to which the path of the optimal investment-*GNP* ratio converges equals the equilibrium investment-*GNP* ratio of the conventional growth model incorporating neither a demand sector nor a foreign sector. Relying upon this and the turnpike property of the ordinary growth model,[5] we propose to approximate the optimal investment-*GNP* ratio of the general model (1)–(4) by solving the following model.

(6) Maximize:

$$W = \int_0^\infty e^{-rt}\{U[(1-s)(1-ucu)\cdot F(K, L; t)] + G(DFR^{*f})\}dt$$

(7) Subject to: $DK = sF(K, L; t) - \delta K$

where $DFR^{*f} = X(t)^f - M(t)^f + j(t)$ and where $M(t)^f$ and $X(t)^f$ are the full employment levels of exports and imports respectively. The two results just noted assure us that the demand sector does not affect the equilibrium level of the optimal investment-*GNP* ratio, and the optimal investment-*GNP* ratio remains in the neighborhood of the equilibrium most of the time. Thus the solution to (6)–(7) should provide a good approximation to the optimal investment-*GNP* ratios. To ensure that the open model is consistent with the two-step solution procedure we defined the full employment level of foreign reserves as FR^f. This represents the potential level of foreign reserves under the conditions of full employment. The full employment level of exports $X(t)^f$ was obtained simply by drawing a straight line through the two highest levels which occurred during the sample period. The full employment level of imports $M(t)^f$ was obtained by substituting full employment level of *cu* into the import equation. We then eliminated equation (2) and made G in (1) a function of the change in the level of foreign reserves, $DFR^{*f} = DFR^f - D\widetilde{FR}^f$. In the long run, it is the trend in FR^{*f} which is important, and since any temporary problem in the balance of payments can be handled by stabilization policies, these modifications seemed to be quite reasonable in order to obtain the first approximation to the optimum.

The second step is concerned with the question of stabilization policy, or the determination of the optimal rate of capacity utilization. Since the demand sector of the economy is highly dependent upon the supply sector we have to incorporate the supply sector into our approximation procedure. To do this we shall use the optimal investment-*GNP* ratio and capital stock generated in the first step. If we substitute

[5] See Sakakibara and Paul Samuelson.

them into the function U and expand it by a Taylor series around the point $ucu=0$, the term $U(C)$ in the social welfare function can be treated as a linear function of ucu. Also, it is reasonable to assume that the functions L, Q and G are quadratic while g and h are linear.[6] Making these additional assumptions, and making the change of variables $\widetilde{ucu}=ucu+k$, where k is a constant, we now have to solve

(8) Maximize:

$$W=\int_0^\infty e^{-rt}[\alpha p^2+\beta\widetilde{ucu}^2+\gamma FR^{*2}]dt$$

(9) Subject to: $Dp=a(\mathrm{t})+b\widetilde{ucu}(\mathrm{t})+cp$

(10) $DFR^*=X(\mathrm{t})-M(\mathrm{t})+j(\mathrm{t})$

With $X(\mathrm{t})$ given exogeneously, and assuming the import equation is of the following form

(11) $M(\mathrm{t})=m_1+m_2cuF(K,L,\mathrm{t})+m_3cu$

equation (10) reduces to

(12) $DFR^*=d(\mathrm{t})+e(\mathrm{t})\widetilde{ucu}$

where $a(\mathrm{t})$, $d(\mathrm{t})$, and $e(\mathrm{t})$, are time varying coefficients and b, c, m_1, m_2, and m_3 are constant.

The import equation (11) is quite reasonable. The term m_3cu can be considered as a proxy for price effects. Price terms in import equations usually turn out to be statistically insignificant. Since capacity utilization terms usually appear significant, cu can be quite a good proxy. Moreover, the cu term can be measuring the effect of queues in domestic production. This can be quantitatively more significant than price effects.

The problem in the second step reduces to the maximization of (8) subject to (9) and (12). The problem is the familiar one of maximizing a quadratic criterion subject to linear system equations. A solution can be derived in feedback form which satisfies the Riccati equations.[7] The feedback solution for $\widetilde{ucu}(\mathrm{t})$ is,

$$\hat{\widetilde{ucu}}(\mathrm{t})=-1/(2\beta)(b,e(\mathrm{t}))(z(\mathrm{t})'+N(t)\begin{bmatrix}p(\mathrm{t})\\FR^*(\mathrm{t})\end{bmatrix}\qquad(13)$$

where $z(\mathrm{t})$ ($z(\mathrm{t})'$ is the transpose of $z(t)$) and $N(\mathrm{t})$ are 1×2 and 2×2 matrices satisfying the following equations.[8]

$$rN-N\begin{bmatrix}c&0\\0&0\end{bmatrix}-\begin{bmatrix}c&0\\0&0\end{bmatrix}N+1/(\beta)\cdot N\begin{bmatrix}b^2&be(\mathrm{t})\\be(\mathrm{t})&e(\mathrm{t})^2\end{bmatrix}N-\begin{bmatrix}\alpha&0\\0&\gamma\end{bmatrix}=0\qquad(14)$$

(15)

$$rz(\mathrm{t})-z(\mathrm{t})\begin{bmatrix}c&0\\0&0\end{bmatrix}+z(\mathrm{t})\begin{bmatrix}b\\e(\mathrm{t})\end{bmatrix}[b\ e(\mathrm{t})]\cdot N(\mathrm{t})-2(a(\mathrm{t}),d(\mathrm{t}))N(\mathrm{t})=0$$

Thus, the optimal solution $\widetilde{ucu}(\mathrm{t})$ is given as a linear function of the current values of state variables $p(\mathrm{t})$ and $FR^*(\mathrm{t})$.

II

In this section we shall apply the model to the *U.S.* economy during 1952–67 to evaluate performance of economic policies. First, using the first step of the two-stage approximations suggested above, we shall evaluate the performance of growth policies.

Key functions of the model in the first stage, production function and imports equation, are estimated by ordinary least squares as follows:[9]

$$\log F(K,L,\mathrm{t})=4.6024+0.263\log K(\mathrm{t})+0.0269\mathrm{t}\qquad(16)$$
$$(27.6)\qquad \bar{R}^2=0.9845\qquad s=0.0132$$

[6] For a treatment of the case where function g is not linear, see N. H. McClamroch et al. and Sakakibara.

[7] See, for example, M. A. Athans and P. Falb, pp. 750–815.

[8] Note that the Riccati equations are time variant because of time varying parameters $a(\mathrm{t})$, $d(\mathrm{t})$, and $e(\mathrm{t})$.

[9] Labor force, L, is assumed to grow at a constant

$$M(t) = -49.0782 + 0.0732F(K, L, t) \atop (6.68)\quad (22.7)$$

$$(17)\qquad +35.3737cu \atop (4.57) \qquad \bar{R}^2 = 0.9721$$

$$s = 1.1814$$

Also, full employment level of exports, $X(t)^f$, is fitted by hand through the two peak levels of exports:

$$(18)\qquad X(t)^f = 18.2e^{(0.0575)t}$$

The social welfare function (6) is specified to:[10]

$$(19)\qquad W = \int_0^\infty e^{-0.1t}A((1-s)F(K, L, t))^\rho + B(TB - \widetilde{TB})^2 dt$$

where TB is the actual trade balance and $\widetilde{TB}$ is the desired level of trade balance. Because of the difficulty involved in introducing international capital transactions formally into the model we have used the trade balance as a proxy for the changes in foreign reserves. A, B, and ρ are constant.

We have first neglected the foreign sector assuming $B=0$ and generated optimal investment-*GNP* ratios for values of $\rho = 0.2$, 0.6, and 0.8 with $A=1.0$. Although the assumption that $\rho=0.2$ seems unrealistic, we have done an experiment with it for reference. A significantly large reduction of ρ by 0.6 caused the equilibrium investment-*GNP* ratio to drop by only 1.4 percentage points. Apparently, the optimal path of the investment-*GNP* ratio is rather insensitive to changes in the value of ρ. All three experiments appear consistent with the turnpike theorem as about 11 to 13 years out of the 15 are spent around the equilibrium investment-*GNP* ratio. Moreover, the investment-*GNP* ratio converges monotonically from above in all three of these cases. Although low values of ρ do not change the equilibrium investment-*GNP* ratio significantly, they substantially cut the high initial values of the investment-*GNP* ratio. In all three cases 1967 constant dollar *GNP* is above $900 billion. This is about $200 billion higher than the actual. Naturally, the optimal investment-*GNP* ratios are 3–5 percentage points higher than the actual values during most periods. The result is at least partly affected by our complete neglect of the trade balance terms.

The next experiment is carried out for three different values of B. We set $B=0$, $B=1$, $B=+\infty$ with $A=1$, $\rho=0.8$ and the rate of embodiment of technological change in capital at 1.2 percent.[11] The results are illustrated in Figure 1. The case where $B=0$ corresponds to the closed economy situation considered earlier. At the opposite extreme is the case where $B=+\infty$; this corresponds to a complete neglect of the consumption term in the utility function. The solutions of these two polar cases bound the solution of the intermediate case, where $B=1.0$. It can be seen that increases in the amount of weight placed on the balance of trade in the utility function lowers the optimal path of the investment-*GNP* ratio. The path of the actual investment-*GNP* ratio is also plotted in the figure. It can be seen to fall near the lower bound of the set of optimal paths.

The introduction of the foreign sector does not necessarily lower the equilibrium investment-*GNP* ratio. For all but the case where $B=+\infty$, the equilibrium investment-*GNP* ratio is around 19.5 per-

rate. Since the ordinary regression produced poor result (negative coefficient on $K(t)$) because of multicollinearity between t and K, we have iterated for values between 0.15 and 0.45 for the coefficient on $K(t)$ and chose the one which gave the highest R^2.

[10] Data are obtained from *Survey of Current Business*. Capital stock data are generated by taking $212.4 billion in 1948 as a starting stock (*SCB*) and by applying estimated depreciation rate of .088.

[11] For theoretical analysis and empirical experiments of embodiment of technological change, see Sakakibara.

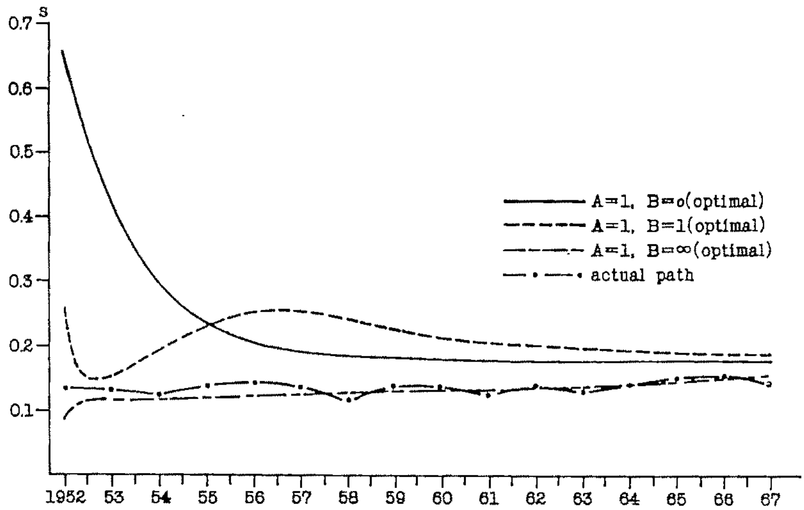


FIGURE 1. U.S. OPTIMAL INVESTMENT GNP RATIOS WITH BALANCE OF PAYMENTS CONSIDERATIONS ($\rho=0.8$)

cent. The turnpike property of the closed economy carries over into the open economy, at least for the one intermediate case considered, but unlike the model with no foreign sector, monotonic convergence is not present in the open economy. The optimum investment-*GNP* ratio does oscillate, particularly in the initial 6 or 7 years. In the last 8–9 years, however, the investment-*GNP* ratio converges to the equilibrium level.

Using equations (17) and (18) the implied trade balance corresponding to each of the three alternative optimal paths was calculated.[12] The first two paths, for which the equilibrium investment-*GNP* ratio is around 19.5 percent, all generate substantial current deficits in the balance of trade although they do decrease as B increases. Thus, the first two paths generate investment-*GNP* ratios which may be too high in the opinion of those who place heavy weight on the trade deficit. They might claim B should be greater than 1.0. But their argument would be misleading in at least three respects. First, equation (17) only approximates the full employment level of imports around the data used to estimate the equation. If *GNP* were to diverge substantially from them (17) would not provide a good approximation. Second, the effect of new investment on both import substitution and exports has been completely neglected. Third, we have also completely neglected the repercussion from the rest of the world for a higher rate of growth and resultant high volumes of imports of the United States. Together, the second and third factors would probably increase the level of exports substantially. This means that equation (18) is much too conservative an estimate of the amount of exports under the cases $B=0$ and 1.0. Therefore, the trade deficits, as we have calculated them, are much too large. In fact, the optimal investment-*GNP* ratios shown might well generate trade surpluses. Since at this point we have no realistic estimates of these effects we are forced to conclude that the optimal path of the investment-*GNP* ratio lies somewhere between the path for $B=1.0$ and $B=+\infty$.

Given these results, it seems safe to say that the actual investment-*GNP* ratio in

[12] We have used the homogeneous version of (17) because of the fact that homogeneous regression provide better long-run estimates in most cases.

the United States has been much too low during the past 15 years, remaining as it has near the lower bound of the optimal paths calculated in this section. Since, however, we have no way to determine the effect of new investment on import substitution and on exports, nor the effects of the other two factors mentioned in the paragraph above, the optimal path for the *U.S.* economy cannot be specified. Nevertheless, it is our guess that an investment-*GNP* ratio of around 16 percent or 18 percent would have been much better than the actual ratio. More particularly, the very low investment-*GNP* ratios experienced during the late 1950's and early 1960's are certainly not justifiable, unless additional terms are included in the social welfare function.

We now want to solve the second stage of the procedure described in the previous section to obtain the optimal path of capacity utilization (*or unemployment rate*) to evaluate the performance of stabilization policies. Since we wish to evaluate the stabilization policies independently from growth policies, we have taken the actual paths of capital accumulation in the past fifteen years as a given constraint. The additional state equation representing price inflation is estimated by ordinary least squares as follows:

$$(20)\qquad Dp = 3.47 - \underset{(3.7)}{0.2312ucu} - \underset{(4.0)}{0.3549p} \qquad \bar{R}^2 = 0.671 \quad s = 0.49$$

Also the actual path of exports is assumed to be a given constraint and the level of foreign reserves, *FR*, is approximated by the accumulated trade surpluses because of the difficulty in introducing international capital transactions into the analysis. The desired level of accumulated trade surpluses is assumed to be 4.0t where t is an index of time. Also the rate of pure time preference is assumed to be 0.1.

The relation between optimal unemployment and excess capacity is given by the estimated equation $\hat{u}(t) = 0.447\widehat{ucu}(t)$. This is a rough approximation to the level of $\hat{u}(t)$ but since we have $\widehat{ucu}(t)$ as a solution we can generate $\hat{u}(t)$ with a more complicated equation, possibly involving lags. In these experiments, however, we have used the result of the simple homogeneous regression indicated above.

Since we do not know the appropriate weights to be put on the parameters of the welfare function (8), we have experimented with various combinations of α, β, and γ.[13] The results are presented in Figure 2 and indicate the high sensitivity of the optimal solution to the values of these parameters. It is interesting to note that the two properties which hold for a simpler model[14]—monotonic convergence to the equilibrium and the turnpike property—hold approximately in this case. They hold only approximately because of an extra state equation which involves the time variant component, the actual amounts of exports, and full employment *GNP*. Specifically for $\alpha = -1.0$, $\beta = -0.2$, and $\gamma = -0.01$, optimal $\hat{u}(t)$ starts from 6.76 percent in 1952 and then approaches the 5.7 percent level gradually from above. With $\alpha = -1.0$, $\beta = -0.8$, and $\gamma = -0.01$, $\hat{u}(t)$ starts from 3.0 percent and then increases to the level of 3.6 percent.

It is also interesting to note that with all

[13] We have not followed the procedure described in Section I exactly because of the arbitrary nature of coefficients in the social welfare function. Instead of approximating $[(1-s(t))(1-ucu(t))K(t)]$ formally through Taylor expansion, we incorporated this in the coefficient β on ucu^2. Mathematically, this procedure is not rigorous enough since approximating a linear function by a quadratic function is difficult. However, since the relative weights between U and ucu^2 are hard to determine and since we have no ideas about size, we have adopted the less formal procedure of adjusting β in this and the following section, hoping that since the linear term in ucu is small the less formal procedure will provide a good approximation.

[14] See Sakakibara.

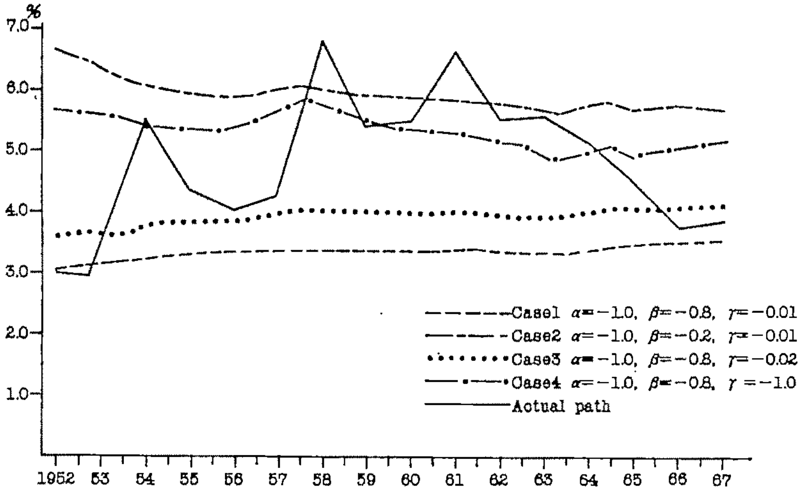


FIGURE 2. OPTIMAL AND ACTUAL UNEMPLOYMENT RATES

combinations of the values of parameters with which we experimented, the deviation of the optimal paths from the equilibria is rather small. The maximum deviation, occurring at the initial period, is never beyond 1.2 percentage points. Thus, we can safely conclude that regardless of the value judgements we might have, it is not wise to generate such erratic movements in the unemployment rate as have occurred in the past. Moreover, the monotonic convergence and turnpike properties of the optimal solution justify fairly vigorous fiscal and monetary policies at the initial stages. The actual path of the unemployment rate is depicted in Figure 2 and we can clearly see that the actual paths are much more bumpy than any of the optimal paths.

It may be useful to show the magnitudes of coefficients in the feedback formula (13) to check the optimum level of reaction of excess capacity to price changes. For $\alpha=-1.0$, $\beta=-0.8$, and $\gamma=-0.02$, the feedback formula is:

$$\widehat{ucu}(t) = \omega_1(t) + 0.31335\dot{p}(t) - 0.06384FR^*(t) \tag{21}$$

In terms of the unemployment rate, (21) becomes

$$\hat{u}(t) = \omega_1(t) + 0.1401\dot{p}(t) - 0.0285FR^*(t) \tag{22}$$

The optimal reaction to a 1 percent price change is as small as a 0.1401 percent change in the unemployment rate. The coefficient on $\dot{p}(t)$ for the case where we put the maximum weight on the price inflation is 0.4655 (for (22)) and even in this most conservative case the optimal reaction rate is very small. Because of the nature of the state equations, it turned out that the two coefficients are time invariant. $\omega_1(t)$ is time variant coefficient, reflecting the time variant nature of (10).

Thus from the above experiments we can safely conclude that regardless of the values of the parameters of the welfare function,[15] the optimal paths of unemployment have to be fairly constant over time and we can have fairly vigorous policies at the initial stages.

[15] As in other sections, the effect of changes in r is yet to be seen. However, from a few experiments we carried out, our conclusions are not changed significantly if r remains in the range between 0.03 and 0.2.

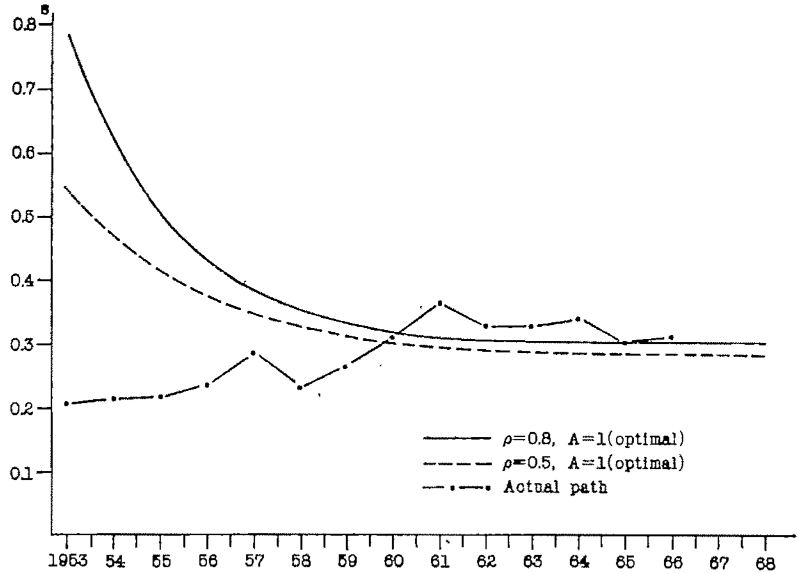


FIGURE 3. OPTIMAL AND ACTUAL INVESTMENT-GNP RATIOS FOR JAPAN

III

In this section we shall apply the same model to evaluate economic policies of Japan between 1953–68. For reasons of space we will just state the conclusions for Japanese growth policy.[16]

The result is presented in Figure 3 for $\rho=0.8$ and $\rho=0.5$.[17] We recognize, of course, that the very high initial optimal rates of investment are not realistic, especially when political and other considerations are taken into account, and probably these should be lowered to around 30–40 percent as practical approximations to the optimal values. Examining the Japanese experience shows that during the decade of the 1950's the actual investment-*GNP* rate was below this optimum but that the initiation in the 1960's of policies intended to foster a high rate of growth raised the actual investment-*GNP* rate until it was about equal to the optimum.

As in the last section, we have applied the feedback formula to evaluate the past demand (stabilization) policies of Japan. There are two substantial institutional differences between the case of Japan as compared to the United States. First, through the lifelong employment system, the full employment of the labor force is assured for the economy, whatever the levels of excess capacity. Unless excess capacity is high enough to threaten a recession, the economy can tolerate a fairly high excess capacity in order to cope with the problem of temporary imbalance of international transactions and price inflation. Second, not being a key currency country and having started with a fairly low level of foreign reserves, Japan had no choice but to put a very high weight on the level of foreign reserves. In the framework of our model, this will increase the cost associated with the deviation of the actual accumulated trade balance from the desired one.

[16] For detailed analyses of this problem see Sakakibara.

[17] Foreign reserves did not turn out to be an effective constraint for growth policy and B was put equal to zero. For details, see Sakakibara.

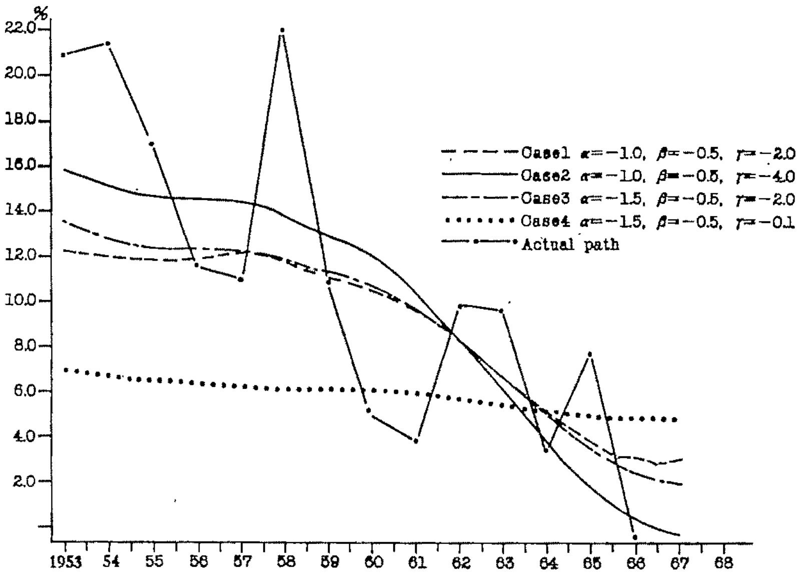


FIGURE 4. OPTIMAL EXCESS CAPACITY RATES (JAPAN)

The dynamic price equation, as before, is estimated by ordinary least squares:

$$(23) \quad \begin{aligned} Dp &= \underset{(13)}{3.6109 - 0.1532ucu} \\ &\quad \underset{(1.8)}{- 0.5463p} \qquad \bar{R}^2 = 0.38 \\ &\qquad\qquad\qquad\quad s = 2.1 \end{aligned}$$

The price index used here is the consumers' price index. Equation (23) is not too good statistically but it may nevertheless represent the economy fairly well. A slight change in the coefficient in (23) will not change the basic nature of the solution we derive. Also we have used the accumulated trade surpluses as a proxy for the level of foreign reserves and the desired level of accumulated trade surpluses is assumed to be 0.35t where t is an index of time. As before actual level of capital accumulation or the full employment *GNP* are given as a constraint to the present problem to evaluate stabilization policies independently of growth policies.

The numerical results with various combinations of α, β, and γ are presented in Figure 4. The first three cases show one very particular pattern, that is, to start from very high excess capacity in 1953 and approach a very low level in 1967. This is a reflection of the fact that the balance-of-payments situation became much easier in the latter half of the period of the simulation. The steepness of the optimal curve is an increasing function of the values of γ, that is, as the weight on the accumulated trade balance increases we should react more sharply to the improvement in the balance-of-payments situation. With $\gamma = -3.0$, $\alpha = -1.0$, and $\beta = -0.5$, the optimal path of excess capacity starts from 14.3 percent and reaches 2.4 percent in 1966. With $\gamma = -4.0$, $\alpha = -1.0$, and $\beta = -0.5$ excess capacity becomes even negative in 1966.[18]

In case 4, we can see if γ is pushed down to -0.1 as compared to $\alpha = -1.5$, we get a fairly smooth curve (from 6.8 percent to 5.0 percent). But this weight is unrealistic

[18] The capacity utilization index published by the Ministry of International Trade and Industry, Japan, takes the level of capacity utilization in 1966 as 1.0. Thus it is possible that *ucu* becomes negative and that is the reason we did not constrain it to be nonnegative in the experiments.

in view of the relative importance of the balance-of-payments problem in Japan.

Note that because of the importance of the balance-of-payments problem and the time-variant nature of the trade functions, the monotonic covergence and turnpike properties of the *U.S.* model does not appear here.

When we compare the actual paths of $ucu(t)$ with the series of optimal paths generated above, we can see that in the 1950's the actual path was much bumpier and sometimes much higher than the optimal one, but in the 1960's the actual and the optimal paths followed each other fairly well. From these observations, we can draw a tentative conclusion that the demand (stabilization) policies in the 1950's were a little too sensitive to the state of economy and too conservative but in the 1960's they were executed fairly well. The good performance of stabilization policies in the 1960's, however, is largely due to the change in the external situation. The balance-of-payments situation, partly due to the high growth rate policies initiated in the early 1960's and partly because of the sustained economic growth in the United States, improved substantially and gradually removed the constraint for stabilization policies. Given the situation, the coincidence of the optimal and the actual is not surprising.

IV

We have attempted to show in this paper that the dynamic optimization model which integrates demand and foreign sectors to an ordinary growth model can be applied to evaluate actual economic policies. Since we can avoid the dichotomy between growth and stabilization policies in an integrated dynamic model presented in this paper, the results should be much more meaningful than ordinary prescriptions given.

Application of the model to the *U.S.* economy shows that the growth policies in the period 1952–67 have been too conservative while the movement of stabilization tools have been too erratic.

The results of the experiments indicate that the Japanese growth policy in the 1960's has been quite successful while that of the 1950's was too conservative. Since the major problem of the stabilization policy, the balance of payments, has been virtually solved by aggressive growth policies in the 1960's, the stabilization policies at that time were fairly easy and produced the good results whereas the policy makers were too sensitive to the changes in the balance of payments in the 1950's.

The same model can be applied to the formulation of future economic policies provided that the estimated relations between economic variables do not change substantially within the planning period.

REFERENCES

K. J. Arrow, "Application of Control Theory to Economic Growth," Institute of Mathematical Studies in Social Science, Stanford Univ. Tech. Rept. 2, July 14, 1967.

M. A. Athans and P. Falb, *Optimal Control,* New York 1966.

N. H. McClamroch et. al., "An Approach to Sub-optimal Feedback Control of Non-Linear Systems," *Int. J. Control,* May 1967, *5,* 425–35.

K. Mera, "A Generalized Aggregative Model for Optimal Growth with Some Emprical Test," *Int. Econ. Rev.,* June 1969, *10,* 149–62.

L. S. Pontryagin, et al., *The Mathematical Theory of Optimal Processes,* New York 1962.

E. Sakakibara, "Optimal Growth and Stabilization Policies," unpublished Ph.D. dissertation, Univ. Mich., June 1969.

P. Samuelson, "A Catenary Turnpike Theorem Involving Consumption and the Golden Rule," *Amer. Econ. Rev.,* June 1965, *55,* 486–96.

K. Shell, ed., *Essays on the Theory of Economic Growth,* Cambridge, Mass. 1967.

The Simple Economics of Incentive Contracting

By J. J. McCall*

This paper presents an elementary economic analysis of three different kinds of procurement contract: the fixed price contract, the fixed price incentive contract, and the cost plus fixed fee contract.[1]

In the fixed price contract, a firm agrees to deliver a particular product for a specified price. If the product's cost is less (more) than the fixed price, the firm incurs a profit (loss). The cost to the government is simply the fixed price. Once the fixed price is established, the government is unaffected by the product's cost—it shares in neither the firm's profits nor losses.

In the fixed price incentive contract, a firm initially establishes a provisional or "target" cost estimate. Target profits are estimated by multiplying the target cost estimate by a specified proportion. In addition, if actual costs are less (more) than the estimated cost, the government shares in the firm's "incentive" profits (losses). The extent of the sharing is indicated in the contract by a fixed proportion which, when multiplied by the profits or losses, represents the firm's share; the remaining portion is the government's share. This contract usually places limitations on both the total amount of profits a firm can earn and on total government costs.

In the cost plus fixed fee contract (*CPFF*), the government pays a firm a fixed percentage of an initial target cost. If actual costs exceed the target cost, the government usually pays the additional expenses. The fixed fee or "profits," however, remain constant.[2]

From an analytical point of view, the fixed price and the cost plus fixed fee contract are both special cases of the fixed price incentive contract. Accordingly, the same model can be used to analyze all three. To avoid needless repetition, only the fixed price incentive contract will be analyzed in detail.

It is frequently alleged and sometimes verified (for certain kinds of contracts) that an awkward tendency to underestimate costs dominates government contracts in general, and military contracts in particular.[3] In their study of cost estimates for French public works, R. Giguet and G. Morlat have devised a purely statistical

* Professor of economics, University of California, Los Angeles. The author is indebted to K. L. Deavers and the referee for their valuable criticisms and suggestions.

[1] The literature on the economics of incentive contracting is quite extensive. See I. N. Fisher, T. Marschak et al., F. T. Moore (1962, 1967), M. J. Peck and F. M. Scherer, J. W. Pratt, and F. M. Scherer (1964a, 1964b). The analysis presented here is implicit in several of these studies, but has never been stated explicitly.

[2] For a more complete description of these contracts, see the studies by Fisher and Moore (1962, 1967).

[3] Neither the tendency to underestimate costs nor the indignant response to underestimates is peculiar to modern times. Roman decision makers faced similar problems in their construction activities: "The young magistrate (Herod, son of Atticus), observing that the town of Troas was indifferently supplied with water, obtained from the munificence of Hadrian three hundred myriads of drachmas (about a hundred thousand pounds) for the construction of a new aqueduct. But in the execution of the work the charge amounted to more than double the estimate, and the officers of the revenue began to murmur, till the generous Atticus silenced their complaints by requesting that he might be permitted to take upon himself the whole additional expense." Edward Gibbon, *The Decline and Fall of the Roman Empire*, Vol. 1, pp. 28–29, abridged by D. M. Low, Washington Square Press, New York, 1962. Needless to say, tranquilizers like Atticus are no longer available to sooth Congressional unrest.

explanation for this phenomenon. One of their basic analytic assumptions is that contractors are trying to obtain unbiased predictions of actual costs. An alternative economic explanation of the procurement process suggests that contractors may not try to be unbiased in estimating costs when this conflicts with their overriding object to maximize expected profits. The economic motivation of the contractor may be the major factor in determining his propensity to over- or under-estimate costs.

When two organizations do business with each other, the profit maximizing behavior of one is symmetrical with the behavior of the other. This symmetry is not present in business-government relations. One reason for its absence is the difference in the criteria each organization attempts to satisfy. The business firm's goal is to maximize profits; given the defense budget, the goal of the defense organization is to maximize national security, a much more ambiguous goal. Perhaps an even more basic factor is the difference in property rights between these two organizations.[4] Within the firm, private property rights in the private sector make it relatively easy to transmit incentives consistent with its objective to maximize profits. Within a government organization, it is much more difficult to transmit incentives consistent with its primary objective.

The analysis indicates several problems associated with the use of incentive contracts. In brief, it will be shown that whenever the sharing proportion is less than unity, i.e., for any contract except the fixed-price variety, incentive contracting induces a perverse divergence between estimated costs and actual costs. This divergence is most perverse for the *CPFF* contract. Relatively efficient firms, i.e., firms whose actual costs of production are low, tend to submit estimated costs (bids) that are higher than actual costs; and relatively inefficient firms, i.e., firms whose actual costs are high, tend to submit estimated costs that are lower than actual costs. If the government chooses among firms on the basis of estimated costs, then it is possible it will select relatively inefficient firms (high actual costs) instead of relatively efficient ones (low actual costs). The probability of selecting a high-cost instead of a low-cost firm is a decreasing function of the sharing proportion. The probability of making an inappropriate choice is highest for the *CPFF* contract where the sharing proportion is zero. To avoid such mistakes, the government's best policy is sometimes the paradoxical one of awarding contracts to firms submitting the *highest* estimated costs (bids). Costs to the government of a given product are *minimized* only if this policy is followed. For this paradoxical policy to be optimal, firms must believe that the contract will be awarded to the lowest bidder. As soon as firms recognize that contracts are being awarded to high bidders, they will substantially alter their bidding behavior. The impossibility of concealing such a policy from contractors, together with its paradoxical character, should convince the reader that *this policy is not being recommended.* It is only mentioned to illustrate the difficulties inherent in the structure of incentive contracts.

In many cases the government awards contracts by some method other than competitive bidding. This alternative selection device has never been clearly specified, contract costs being targeted after the firm is selected. The foregoing remarks will be relevant if the choice of a firm is conditioned by anticipated target costs.

Certainly, the government wishes to design contracts that attract efficient firms, and at the same time adhere to con-

[4] See Armen Alchian.

ventional bidding practices. Incentive-type contracts do not permit both of these goals to be achieved simultaneously. The overriding goal of the government is to attract efficient firms. This can be accomplished either by altering the structure of incentive contracts or by imposing a set of bizarre bidding rules. The first alternative is clearly preferable.

I. *An Economic Model of Procurement*

This section presents a simple economic model of procurement. The purpose of the model is to understand and predict the behavior of firms when they are confronted with alternative types of government contracts.

The firm is assumed to maximize expected profits. This assumption neglects the influence of variance (risk) and higher order moments of the probability distribution of profits.[5] The cost of producing a particular item is assumed to be a random variable with a known probability distribution. Similarly, the receipts obtainable from selling a product using similar resources in the private sector is also a random variable with a known probability distribution. The firm is assumed to have no knowledge of other firm's costs and, therefore, is ignorant of the bids competitors submit. To clarify the presentation, the analysis first assumes that both the receipts from private sector production and the production costs are known. Then it is assumed that only the production cost is a random variable. The adjustment required when both receipts and production costs are random variables is presented in a footnote.

To begin with, it will be assumed that the firm submits a single bid (target cost) on which all incentives are based. Since the firm is assumed to have the alternative of producing in the private sector, it will never submit a target cost (bid) less than some minimum amount, this figure being determined by the profits expected in the private sector. Furthermore, being unaware of the production costs of its competitors, the existence of competitors will deter the firm from submitting bids that are much above this minimum.[6] In the model, target costs or bids are calculated to equate profits from selling to the government and to the private sector.[7]

In presenting the model, the following notation will be employed:

C_1 = A random variable denoting the actual cost of producing the product or service.

ϕ = The probability density function associated with C_1.

C_2 = The cost of the product to the government.

R = A random variable denoting receipts from producing in the private sector.

p = Target cost or bid the firm submits.

π_1 = A random variable denoting profits from selling in the private sector.

π_2 = A random variable denoting profits from selling to the government.

α = The target profit rate, a proportion that when multiplied by target

[5] In Appendix 1 attitudes towards risk are included in the analysis. Also see Fisher and G. R. Hall.

[6] The relationship between the firm's bid and its uncertainty regarding the bids of competitors is examined in Appendix 2.

[7] Typically, firms that produce products for the government establish a government division separate from the rest of the organization. Under these circumstances, it seems likely that the performance of the government division will be judged relative to the performance of the other divisions. It furthermore seems likely that the manager of the government division will try to match the rate of return on capital of the other divisions. The importance of this influence will, of course, depend on how easily resources can be transferred from the government sector to the private sector. In the short run, firms with highly specialized resources will be relatively less responsive to private market alternatives. Nevertheless, this study assumes that, on the average, the influence of private market alternatives on individual contracts is both measurable and predictable.

costs yields target "profits"; $\alpha>0$.

αp = Target profits.[8]

β = The incentive profit rate, the fraction of the difference between p and C_1 that accrues to the firm; $0\leq\beta\leq 1$.

$\bar{C}_2$ = The maximum price the government will pay for the product; the firm absorbs costs of production in excess of this maximum.

$\bar{\pi}_2$ = The maximum profits allowed on a particular government contract; profits in excess of the maximum are returned to the government.

The following relations are immediate consequences of these definitions:

$$\pi_1 = R - C_1 \tag{1}$$

$$\pi_2 = C_2 - C_1 \tag{2}$$

The Fixed Price Incentive Contract

Given that the firm has the alternative of producing in the private sector, the competitive bids for government contracts will tend to equate profits from the government contract with alternative profits available in the private sector. Assume for the moment that both firm profits and government costs are unrestricted by the fixed price incentive contract, maximum profits $\bar{\pi}_2$ and maximum costs $\bar{C}_2$ are both set equal to infinity. The profits, π_2, derivable from the government incentive contract are given by

$$\pi_2 = \alpha p + \beta(p - C_1), \qquad \alpha > 0, \quad 0 < \beta < 1, \tag{3}$$

where p denotes the firm's bid or target cost, αp represents target profits, C_1, a random variable, denotes the costs of production, and β is the proportion of the algebraic difference between p and C_1 that accrues to the firm.

Alternatively, the profits, π_1, obtainable in the private sector are

$$\pi_1 = \begin{cases} R - C_1, & \text{when } R \geq C_1 \\ 0, & \text{when } R < C_1, \end{cases} \tag{4}$$

where R and C_1 denote, respectively, receipts and cost of production. If the firm's profits in the private sector are negative ($R<C_1$), in the long run the firm will always move to another industry where its profits are zero.

The bid or target cost the firm submits is calculated by equating π_1 and π_2, and solving for p. This yields the relation

$$p = \begin{cases} \dfrac{R - (1-\beta)C_1}{\alpha + \beta}, & R \geq C_1 \\[2ex] \dfrac{\beta C_1}{\alpha + \beta}, & R < C_1 \end{cases} \tag{5}$$

Production Costs Known

When C_1 is known, firms whose private sector receipts exceed the production cost ($R\geq C_1$) are said to be efficient; similarly, firms whose production costs exceed private sector receipts ($C_1>R$) are said to be inefficient. Under conditions of certainty and in the absence of government contracts, inefficient firms will move to other industries. In the presence of incentive contracts, however, it may pay inefficient firms to remain in the same industry and offer all of their output to the government.[9]

The relation (5) between bid and actual production costs is shown in Figure 1.[10] For efficient firms, the target cost is a decreasing linear function of actual costs,

[8] The term target profits is commonly used in the literature on incentive contracts. It affects economic profits, but should not be confused with them.

[9] This assumes that the zero profits guaranteed by the incentive contract are superior to any that could be made in a private industry. As usual, normal profits are included in costs.

[10] In the fixed price contract, $\beta=1$ and $\alpha=0$ and there is no divergence between the C_2 and p curves of Figure 1. In the *CPFF* contract, $\beta=0$ and hence for a given α the divergence between these two curves is maximal.

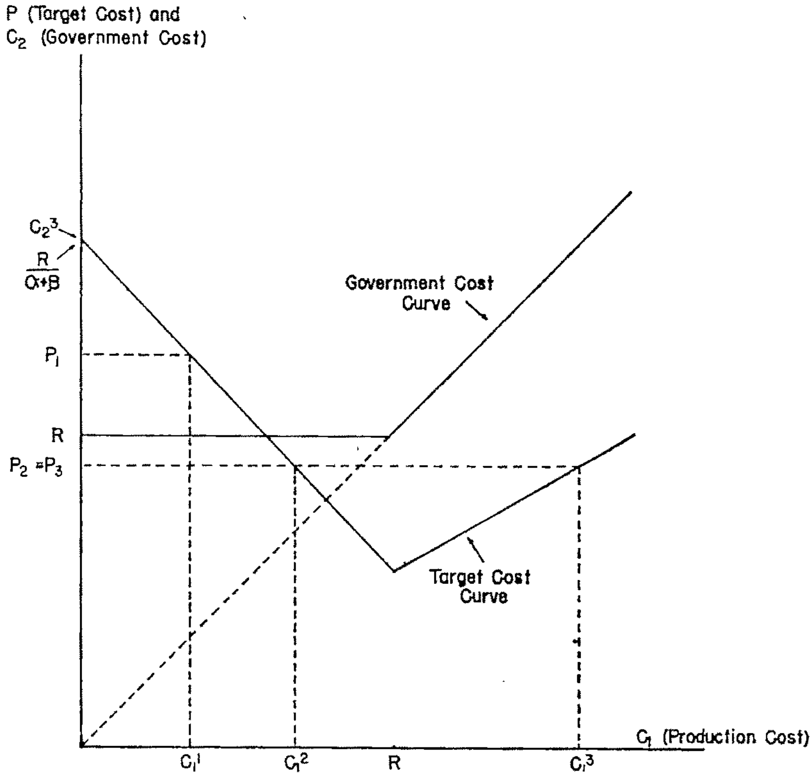


FIGURE 1. TARGET COSTS, P, AND GOVERNMENT COSTS, C_2 AS A FUNCTION OF COSTS OF PRODUCTION, C_1 ($\alpha>0, \beta>0$).

where the rate of decrease m_1 is

$$m_1 = \frac{1-\beta}{\alpha+\beta} \tag{6}$$

For inefficient firms, the bid is an increasing function of actual costs, where the rate of increase m_2 is

$$m_2 = \frac{\beta}{\alpha+\beta}, \qquad m_2 < 1 \tag{7}$$

Firms whose actual costs are less than $R/(1+\alpha)$ will submit bids greater than actual costs—an underrun will occur. These firms are efficient. Firms whose actual costs are greater than $R/(1+\alpha)$ will submit bids less than their actual costs of production—an overrun will occur. This latter group includes efficient firms whose costs lie in the interval $R/(1+\alpha), R$ and all inefficient firms. Notice that efficiency and inefficiency may be equivalently defined in terms of underrun and overrun, respectively, only when the proportion α is set equal to zero.

The relation between government costs and actual costs is given by

$$C_2 = \begin{matrix} R, & \text{when } C_1 < R \\ C_1, & \text{when } C_1 > R \end{matrix} \tag{8}$$

This is obtained by substituting equation (5) into the expression for government costs, $C_2=(1+\alpha)\ p+(1-\beta)\ (C_1-p)$. The C_2 curve in Figure 1 represents this relation. Government costs are constant and equal to R as long as contracts go to efficient firms. Government costs are equal to actual costs whenever the government does business with inefficient firms.

As an example, suppose the government receives bids from three firms labeled 1, 2, and 3. The production costs for these firms are respectively, C_1^1, C_1^2, and C_1^3; the corresponding bids are, respectively, p_1, p_2, and p_3, where $p_2=p_3$. Consequently, the government will reject firm 1 and be indifferent between firms 2 and 3; whereas on the basis of costs (C_2 curve), the government should reject firm 3 and be indifferent between firms 1 and 2. Because of the perverse divergence between bid price and government cost, the government cannot distinguish between bids from an efficient firm (p_2) and from an inefficient firm (p_3); and even when the bids are not identical, it may select the inefficient firm (firm 3) instead of the efficient one (firm 1).

The presence of maximum cost and maximum profit restrictions necessitates a slight modification of the analysis. The first restriction affects only inefficient firms and is denoted by

$$C_2 \leq \bar{C}_2, \tag{9}$$

where the maximum cost to the government is

$$\alpha p + C_1 + \beta(p - C_1) = \bar{C}_2 \tag{10}$$

From equation (5) the bid submitted by the inefficient firm is

$$p = \frac{\beta C_1}{\alpha + \beta}, \qquad C_1 > R \tag{11}$$

This assumes that the maximum cost $\bar{C}_2$ exceeds R; otherwise no bids will be submitted. Solving equations (10) and (11) for p and C_1 gives

$$p = \frac{\beta \bar{C}_2}{\alpha + \beta} \tag{12}$$

$$C_1 = \bar{C}_2, \tag{13}$$

which represent, respectively, the maximum bid the inefficient firm submitted and the actual production costs. Any firm whose costs exceed $\bar{C}_2$ will not submit a bid.

The second restriction, a specified maximum profit, only affects efficient firms. The restriction requires that the firm's profits always be less than

$$\alpha p + \beta(p - C_1) = \bar{\pi}_2 \tag{14}$$

But efficient firms will always submit bids of amount

$$p = \frac{R - (1 - \beta)C_1}{\alpha + \beta} \tag{15}$$

Solving equations (14) and (15) for p and C_1 yields

$$p = \frac{\bar{\pi}_2 + \beta(R - \bar{\pi}_2)}{\alpha + \beta} \tag{16}$$

$$C_1 = R - \bar{\pi}_2 \tag{17}$$

The solution for p represents the maximum bid submitted by an efficient firm. If the firm's production costs are less than $R-\bar{\pi}_2$, it will not submit a bid to the government but will instead produce in the private sector.

From equations (12) and (16), the maximum bid by an inefficient firm is less than the maximum bid by an efficient firm whenever the following inequality holds:

$$R > \bar{C}_2 - \bar{\pi}_2\left(\frac{1 - \beta}{\beta}\right) \tag{18}$$

It would appear that this inequality is satisfied for a great many fixed price incentive contracts;[11] and the inequality implies the following paradoxical result. The government's optimal policy for the fixed price incentive contract is to accept the highest bid, provided it is high

[11] In the typical example presented on page 6 of Moore's study (1962), $\bar{C}_2=\$250{,}000$, $\bar{\pi}_2=\$30{,}000$, $\beta=.2$. By equation (19), the maximum bid by an inefficient firm is smaller than the maximum bid by an efficient firm if $R>\$130{,}000$. But the actual production costs were equal to \$185,000, so it does not seem unlikely that receipts from production in the private market would exceed \$130,000.

enough.[12] Any bid greater than $(\beta C_2)/(\alpha+\beta)$ comes from an efficient firm. The maximum cost restriction guarantees that no inefficient firm will submit so high a bid. The government, therefore, can be certain that for bids in this range its costs will be R. For any bid less than $(\beta C_2)/(\alpha+\beta)$, the government cannot be certain whether it is dealing with an efficient firm incurring costs equal to R or with an inefficient firm incurring costs greater than R.[13]

Production Costs Not Known with Certainty

Several modifications are necessary when production costs are not known with certainty.[14] The bid a firm submits is calculated to equate expected profits in the private sector with those in the government sector. Expected profits in the private and government sectors are denoted, respectively, by

$$E\pi_1 = \operatorname{Max}(R - EC_1, 0) \tag{19}$$

and

$$E\pi_2 = \alpha p + \beta(p - EC_1), \qquad \alpha > 0, \quad 0 < \beta < 1 \tag{20}$$

Firms in the private sector are assumed to switch industries whenever expected profits are negative.[15] Equating (19) and (20), and solving for p yields

$$p = \begin{cases} \dfrac{R - (1-\beta)EC_1}{\alpha+\beta}, & R \geq EC_1 \\ \dfrac{\beta EC_1}{\alpha+\beta}, & R < EC_1 \end{cases} \tag{21}$$

The expected costs to the government when it selects a firm characterized by a probability distribution ϕ_i are given by

$$E(C_2) = R\Phi_i(R) + \int_{R\Phi_i(R)}^{\infty} C_1\phi_i(C_1)dC_1{}^{16} \tag{22}$$

where

$$\begin{aligned} \Phi_i(R) &= 0 \qquad \text{when } R < E[C_1] \\ &= 1 \qquad \text{when } R > E[C_1] \end{aligned}$$

and $\phi_i(C_1)$ is the probability density function of the random variable, C_1, for the ith firm.

A probability distribution now characterizes each firm in the competition. The probability that the production costs are greater than market price will be large for

[12] This policy is not being recommended as noted earlier in this discussion.

[13] If the government chooses firms by some method other than competitive bidding, presumably, the firm so chosen retains the option of producing in the private sector. Consequently, if the government chooses an efficient firm, the target cost should be greater than, or equal to, the bid that this firm would present in competitive bidding. Indeed, in the face of hard government bargaining the two should be equal. And this should also be true of inefficient firms. Even if the government utilizes a different selection device, actual costs will continue to exceed target costs for inefficient firms, whereas target costs will exceed actual costs only for efficient firms. In addition, all of the preceding results are applicable in this case, depending on how much anticipated target costs influence the government's choice of firm.

[14] The uncertainty of production costs in the private sector usually differs from production uncertainty in the government sector. These differences could be treated within this model, but for simplicity the two uncertainties are assumed to be identical.

[15] Of course, firms may (and do) experience losses in the private sector. In this model, however, expected profits are always nonnegative.

[16] When both C_1 and R are random variables, the firm's bid and expected government costs are, respectively,

$$p = \begin{cases} \dfrac{ER - (1-\beta)EC_1}{\alpha+\beta}, & ER \geq EC_1 \\ \dfrac{\beta EC_1}{\alpha+\beta}, & ER < EC_1 \end{cases}$$

and

$$EC_2 = \iint_{R-C_1\geq 0} R\Psi(R_1C_1)dRdC_1 + \iint_{R-C_1<0} C_1\Psi(R_1C_1)dRdC_1$$

where ER is the expected value of private market receipts and $\Psi(R_1C_1)$ is the joint probability density function of R and C_1.

inefficient firms and small for efficient ones. The relation between the bid and these different probability distributions and the relation between government costs and these probability distributions have essentially the same form as the relations shown in Figure 1. Whereas before the government was unable to discriminate between low- and high-cost firms, it is now unable to discriminate between firms with efficient probability distributions and those with inefficient probability distributions.

II. *Conclusions*

The incentive contracts the government currently uses appear to have a serious flaw. Because of the peculiar sharing rule that characterizes these contracts, it is difficult for the government to distinguish between high- and low-cost firms on the basis of the submitted bids or target costs. The fact that efficient or low-cost firms must share some of their profits—difference between target costs and actual costs—with the government induces them to submit cost estimates higher than expected costs. On the other hand, inefficient or high-cost firms are permitted to share some of their losses with the government and, consequently, submit cost estimates that are less than expected costs. The problem of discriminating between efficient and inefficient firms becomes less serious as the sharing proportion rises, where this proportion is the fraction of profits (losses) the firms capture. Only when the sharing proportion is set equal to unity—a fixed price contract— are target costs or bids unambiguous indicators of the firm's expected costs.

Some of the previous analyses of cost estimation in government contracts has been primarily statistical. These statistical studies assume that contractors seek unbiased predictions of actual costs and they have been partially successful in explaining the tendency to underestimate costs. In this study, an alternative *economic* explanation of the procurement process suggests that contractors may bias their cost estimates to conform with their overriding objective to maximize profits.

Appendix 1

This Appendix calculates a lower bound on the bid a vendor submits for an incentive type contract. If the contracts are let under competitive bidding the bids will tend to equal this lower bound. The vendor is assumed to have a constant attitude toward risk; that is, the risk premium required for a particular risk is independent of the firm's wealth. The firm may have constant risk aversion, constant risk indifference, or constant risk preference. Bids that firms submit are conditioned by the alternative opportunities available to them. In particular, it is assumed that the bid is calculated so that the expected utility of the incentive contract is no less than the expected utility of the best alternative opportunity. A lower bound on the bid is obtained by equating these expected utilities. The cost, C, of producing for the incentive contract is a random variable with an arbitrary probability distribution. Similarly, the profits, Π, obtainable on the best alternative opportunity are also a random variable with an arbitrary probability distribution.

Constant Risk Aversion

When the firm has constant risk aversion its utility function with respect to profits, Π, may be denoted by

$$u_1(\Pi) = a_1 - b_1 e^{-\rho\Pi}, \qquad > 0, \tag{1}$$

where ρ measures risk aversion, i.e., the risk aversion function is simply[17]

$$\rho(\Pi) = -\frac{u_1''}{u_1'} = \rho \tag{2}$$

Profits, Π_2, from the incentive are

$$\Pi_2 = \alpha p + \beta(p - C), \qquad 0 \leq \alpha < 1, \quad 0 < \beta < 1, \tag{3}$$

[17] See McCall and Pratt.

where α is the target proportion, p is the bid or target cost, β is the sharing proportion, and C denotes the actual costs. If Π_1 is opportunity profits, then both Π_1 and Π_2 are random variables with arbitrary cumulative distribution functions (c.d.f.), say, F_1 and F_2, respectively.

A lower bound on the bid, p, submitted by the firm is obtained by equating the expected value of $u_1(\Pi_1)$ with the expected value of $u_1(\Pi_2)$ and solving for p, where the first expectation is with respect to Π_1 and the second expectation is with respect to C. Symbolically,

$$(4)\quad Eu_1(\Pi_1) = a_1 - b_1E[e^{-\rho\Pi_1}],$$

$$(5)\quad Eu_1(\Pi_2) = a_1 - b_1E[e^{-\rho(\alpha p+\beta(p-C))}]$$

Equating $Eu_1(\Pi_1)$ and $Eu_1(\Pi_2)$ and solving for p yields

$$(6)\quad p = \frac{\ln \Psi_2(\rho\beta) - \ln \Psi_1(-\rho)}{\rho(\alpha+\beta)}$$

where Ψ_1 (t) denotes the moment generating function of Π_1 and Ψ_2 (t) denotes the moment generating function of C_2. For example, if Π_1 and C are distributed normally with mean and variance, m_1, σ_1^2, and m_2, σ_1^2, respectively, then

$$(7)\quad \Psi_1(-\rho) = \exp[-\rho m_1 + (1/2)\rho^2\sigma_1^2]$$

$$(8)\quad \Psi_2(\rho\beta) = \exp[\rho\beta m_2 + (1/2)\rho^2\beta^2\sigma_2^2]$$

It follows that[18]

$$(9)\quad p = \begin{cases} \dfrac{m_1+\beta m_2+\frac{1}{2}\rho[\beta^2\sigma_2^2-\sigma_1^2]}{\alpha+\beta}, & m_1>0 \\[2ex] \dfrac{\beta m_2+\frac{1}{2}\rho[\beta^2\sigma_2^2-\sigma_1^2]}{\alpha+\beta}, & m_1<0 \end{cases}$$

where $\frac{1}{2}\rho(\beta^2\sigma_2^2-\sigma_1^2)$ can be interpreted as the risk premium. Notice that when $\beta=\sigma_1/\sigma_2$, the risk on the incentive contract is identical to that of its alternative opportunity and the risk premium is zero. The risk premium is positive when $\beta>\sigma_1/\sigma_2$ and negative when $\beta<\sigma_1/\sigma_2$.

Also, assuming $m_1>0$, the bid for a fixed-price contract ($\beta=1$) with $\alpha>0$ is given by

$$(10)\quad p = m_1 + m_2 + \tfrac{1}{2}\rho(\sigma_2^2 - \sigma_1^2)$$

Similarly, the bid for a cost plus fixed fee contract ($\beta=0$) with $\alpha>0$ is given by

$$(11)\quad p = \frac{m_1 - \frac{1}{2}\rho\sigma_1^2}{\alpha},$$

which, of course, may be negative.

Constant Risk Indifference

For constant risk indifference,

$$(12)\quad u_2(\Pi) = a_2 + b_2\Pi$$

and

$$(13)\quad Eu_2(\Pi) = a_2 + b_2E\Pi$$

It follows that a lower bound on the bid the firm will submit is

$$(14)\quad p = \begin{cases} \dfrac{E\Pi_1 + \beta EC}{\alpha+\beta}, & E\Pi_1 > 0 \\[2ex] \dfrac{\beta EC}{\alpha+\beta}, & E\Pi_1 < 0 \end{cases}$$

Constant Risk Preference

Finally, if the firm's utility function exhibits constant risk preference,

$$(15)\quad u_3(\Pi) = a_3 + b_3\exp(\rho\Pi), \qquad \rho > 0$$

$$(16)\quad Eu_3(\Pi_1) = a_3 + b_3E[\exp(\rho\Pi_1)]$$

$$(17)\quad Eu_3(\Pi_2) = a_3 + b_3E\{\exp[\rho(\alpha p + \beta(p - C))]\},$$

and

$$(18)\quad p = \frac{\ln \Psi_1(\rho) - \ln \Psi_2(\rho\beta)}{(\alpha+\beta)}$$

For example, if Π_1 and C are distributed normally as before, then

[18] It is assumed that if expected alternative profits are negative, the firm will not produce and will reduce its expected profits to zero.

$$(19)\qquad \Psi_1(\rho) = \exp[\rho m_1 + (1/2)\rho^2\sigma_1^2]$$

and

$$(20)\qquad \Psi_2(\rho\beta) = \exp[-\rho\beta m_2 + (1/2)\rho^2\beta^2\sigma_2^2]$$

It follows that a lower bound on the bid this risk preferent firm submitted is given by

$$(21)\qquad p = \begin{cases} \dfrac{m_1+\beta m_2+(1/2)\rho^2\sigma_1^2-\beta^2\sigma_2^2}{\alpha+\beta} & m_1>0, \\[2ex] \dfrac{\beta m_2+(1/2)\rho^2\sigma_1^2-\beta^2\sigma_2^2}{\alpha+\beta}, & m_1<0 \end{cases}$$

Appendix 2

Optimal Bidding When the Firm Possesses Probabilistic Information Regarding Competitors' Bids

In Section I it was assumed that the firm had no knowledge of other firms' costs and was, therefore, ignorant of their bids. The case is now considered where firms have probabilistic information on competitors' bids. Let $\Phi_n(\xi)$ denote the probability that the bids of the n competitive firms are all greater than ξ. If the bids are stochastically independent, then

$$(1)\qquad \Phi_n(\xi) = \prod_{i=1}^{n} \Phi_i(\xi),$$

where $\Phi_i(\xi)$ is the probability that the ith competitor's bid exceeds ξ.

Assuming the contract is awarded to the firm submitting the lowest bid, the expected profits derived from a bid of p are given by

$$(2)\qquad E\Pi(p) = \Pi_1[1 - \Phi_n(p)] + \Pi_2\Phi_n(p),$$

where Π_1 and Π_2 are defined as in Section I. The firm then chooses p_2 to maximize expected profits, $E\Pi(p)$, $\bar{p}\leq p<\infty$ where $\bar{p}$ denotes the minimum bid the firm would ever submit.

REFERENCES

A. Alchian, "Electrical Equipment Collusion: Why and How," unpublished, Apr. 1961.

K. L. Deavers and J. J. McCall, *Notes on Incentive Contracting*, The RAND Corporation, RM-5019-PR, Sept. 1966.

——— and ———, *An Analysis of Procurement and Product Improvement Decisions*, The RAND Corporation, RM-3859-PR, Dec. 1963.

I. N. Fisher, "An Evaluation of Incentive Contracting Experience," *Nav. Res. Log. Quart.*, Mar. 1969, *16*, 63–83.

——— and G. R. Hall, "Risk and Corporate Rates of Return," *Quart. J. Econ.*, Feb. 1969, *83*, 79–92.

R. Giguet and G. Morlat, *The Causes of Systematic Error in the Cost Estimates of Public Works*, RAND Translation T-76, translation by W. W. Taylor, rev. Mar. 24, 1958.

G. R. Hall and R. E. Johnson, *Aircraft Co-Production and Procurement Strategy*, The RAND Corporation, R-450-PR, May 1967.

T. Marschak, T. K. Glennan, Jr., and R. Summers, *Strategy for R and D: Studies in the Microeconomics of Development*, New York 1967.

J. J. McCall, "Competitive Production for Constant Risk Utility Functions," *Rev. Econ. Stud.*, Apr. 1967, *34*, 417–420.

F. T. Moore, *Military Procurement and Contracting: An Economic Analysis*, The RAND Corporation, RM-2948-PR, June 1962.

———, "Incentive Contracts," in S. Enke, ed., *Defense Economics*, Englewood Cliffs 1967, ch. 12.

M. J. Peck and F. M. Scherer, *The Weapons Acquisition Process; An Economic Analysis*, Boston 1962.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan.-Apr. 1964, *32*, 122–36.

F. M. Scherer, (1964a) "The Theory of Contractual Incentives for Cost Reduction," *Quart. J. Econ.*, May 1964, *278*, 257–280.

———, (1964b) *The Weapons Acquisition Process: Economic Incentives*, Boston 1964.

Job Search, the Duration of Unemployment, and the Phillips Curve

By DALE T. MORTENSEN*

This paper presents a simultaneous equations model that describes the dynamic behavior of money wages and unemployment in an essentially competitive labor market. The model explains the existence of a positive equilibrium level of unemployment, and provides for an inverse relation between the rate of change in money wages and the unemployment ratio. An important result is that the equilibrium unemployment ratio is found to be independent of the inflation rate.

The market is one in which heterogeneous but substitutable labor services are sold. The suppliers of these services do not know the wage offered or the skills required by any particular employer although they are assumed to know the structure and distribution of wage offers. The numbers of participants on both the supply and the demand sides of the market are large so that no one of them has an appreciable influence on the general levels of wages. The decisions faced by both suppliers and employers of services are analyzed. The results of the analyses provide the basis for an aggregate theory of wage and employment adjustment.

One of the aggregate relationships derived can be interpreted as a Phillips curve. In particular, the difference between the proportional rate of change in money wages and the rate anticipated by employers is shown to be inversely related both to the unemployment ratio and to its proportional rate of change. Thus, any adjustment in the anticipated rate of wage inflation shifts the Phillips curve. Moreover, the "equilibrium" or "natural" rate of unemployment, that ratio which will be maintained in the face of a fully anticipated inflation rate, is independent of the inflation rate. This ratio equals the steady state proportion of the labor force who are in the process of searching for an acceptable job. The proportion is positive 1) because new entrants continuously flow into the market and 2) because some time spent searching is rational when information about offers and openings is imperfect.

Several authors have derived similar relationships recently, including Edmund S. Phelps (1968a), Robert E. Lucas and Leonard A. Rapping (1969b), and Mortensen. The assumption of imperfect information is common to all of these analyses. However, the Phillips relation arises for somewhat different reasons in each case.[1] It arises here for yet another. During periods of excess demand for (supply of) labor, each employer simultaneously attempts to raise (lower) his own money wage offer relative to the general level and to lower (raise) the level of skill he requires of new employees. Thus, during

* Assistant professor of economics at Northwestern University. This research was supported by National Science Foundation grant GS2190 for the investigation of short-run employment decisions.

[1] These and other related papers are discussed in a recent survey article by Phelps (1969). All of the papers mentioned are included in Phelps et al.

periods of wage inflation in excess of that anticipated, the level of unemployment falls rapidly. Conversely, when employers wish to accumulate employees at a less rapid rate, wages are bid down relatively and skill requirements increase.

One clear deficiency of this paper is the treatment of the participation decision. In particular, the individual's decision to participate in the labor force is assumed to be insensitive to the wage available. This assumption clearly ignores the work-leisure tradeoff. Lucas and Rapping (1969b) have recently provided a fresh analysis of this choice problem in a dynamic context, but they ignore the search problem. A synthesis of their theory of participation and the theory of job search developed here could provide a richer and more general explanation of labor market phenomena.

The remainder of the paper is divided into four sections. In the first, the process of search is described and the choice problem faced by the participant searching for a job is formulated. The problem essentially involves choosing an "acceptance" wage which, together with the skill of the participant, determines the expected length of time spent searching for an acceptable job. In Section II the relationship between the expected duration of search unemployment and the acceptance wage is investigated. In addition, the implications of the decision for the aggregate long-run level of search unemployment are derived. Demand considerations are introduced in the third section. Moreover, a complete model of labor market dynamics is presented. One of the relationships of this model is interpreted as a Phillips relation. In the final section the unemployment ratio derived in Section II is shown to be the only one consistent with equilibrium in the context of the model. Other aggregate implications of the model are also summarized.

I. *A Model of Job Search*

Consider an unemployed worker who is participating in the labor force in the sense that he is seeking a job. Assume that his information concerning the nature of existing vacancies and offers is imperfect. In particular, he neither knows the skills required for any particular opening nor the wage offered to fill the vacancy. The only information possessed by the participant is the nature of the frequency distribution of all offers and the structure of wage offers across skill levels. To determine whether his own skills meet the requirements for a particular vacancy and to ascertain the wage offered to fill the opening, he must search it. The process of search is one of sequential sampling from the population of vacancies. Given that the participant does not know the relevant properties of any one of the vacancies, the sampling can be viewed as random.

Search requires time. To simplify the analysis, we assume that one and only one vacancy can be searched within a period of specified length. Suppose that offers must be answered as they are discovered. Under these conditions, the participant must determine prior to each vacancy sampled the terms which he would be willing to accept.[2]

The economic problem faced by the participant concerns the choice of the wage he would be willing to accept. On the one hand, the higher he sets his acceptance wage, the longer he can expect to search before finding an offer acceptable to him. On the other, the higher the acceptance wage, the higher the expected wage once employed. The problem is essentially one of optimal investment in human capital. As we shall show, the best choice of the

[2] Essentially these are the assumptions made by George J. Stigler (1962) in his pioneering paper on the theory of job search. The model presented here is a formalization of that suggested in his paper.

acceptance wage is that which equates at the margin the value of time spent searching to the present value of the future benefits attributable to search.

For simplicity we assume that skills can be measured along a single dimension.[3] Let x denote the minimum level of skill allowed by a particular employer and let x^o denote the skill level of the participant in question. In addition, let y be the relative wage offered by the employer; i.e., the money wage offer divided by the average money wage offer in the market. Let y^o denote the relative wage acceptable to the participant in question. The participant becomes employed after searching an employment opportunity with characteristics (x, y) if, and only if, $x \leq x^o$ and $y \geq y^o$.

We assume that each opening requiring the same skill offers the same wage and that the relative wage offered is higher the more highly skilled the applicant must be. In other words,

$$(1) \qquad y = y(x); \qquad y'(x) > 0$$

For any particular vacancy, then, the wage offer is known once the skill requirements are known. In addition, the nature of the frequency distribution of relative wage offers reflects the nature of the distribution of skill requirements over all vacancies and the nature of the function $y(x)$. Assume that the distribution of all relative wage offers can be described by a continuous density function, $f(y)$, with the following properties:

$$(2) \quad \begin{aligned} &(a) \quad f(y) > 0 \quad \text{for all } 1 > a < y < b > 1 \\ &(b) \quad f(y) = 0 \quad \text{elsewhere} \\ &(c)^4 \quad \int_a^b yf(y)dy = \int_a^b f(y)dy = 1 \end{aligned}$$

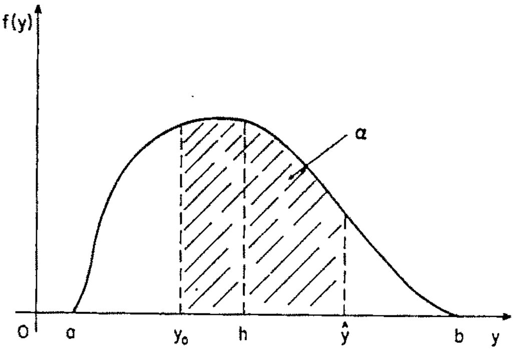


FIGURE 1

Given the assumptions embodied in equation (1), the participant is qualified for all openings which offer a relative wage equal to $y(x^o)$ or less. If we let $\hat{y}$ denote this wage offer, the probability that a randomly selected vacancy is one for which the participant is qualified and is also one which is acceptable to him is given by the following equation:

$$(3) \quad \alpha = Pr(y^o \leq y \leq \hat{y}) = \int_{y^o}^{\hat{y}} f(y)dy$$

The expected value of the relative wage offer received given that the participant is qualified and given that the offer is acceptable is derived in the following manner:

$$(4) \quad h = E(y \mid y^o \leq y \leq \hat{y}) = \frac{\int_{y^o}^{\hat{y}} yf(y)dy}{\int_{y^o}^{\hat{y}} f(y)dy}$$

In Figure 1 we illustrate a possible density function. The probability that a randomly sampled vacancy is open to the participant and acceptable, α, is the shaded area under the curve. The expected relative wage given that the participant qualifies and that the wage offer is acceptable is the mean of the shaded area.

[3] The analysis can be generalized to account for all factors which are important to employers by allowing x to be a vector. Most of the analysis presented in this section holds in the general case.

[4] The mean of the relative wage distribution is unity, provided that each relative wage is constructed by dividing the money wage by the mean of the money wage distribution.

Since the participant samples one vacancy per period, the probability of becoming employed per period is α and the number of periods the participant can expect to search before finding an acceptable job offer equals $1/\alpha$. The expected duration of search is longer the higher the acceptance wage is since

$$\frac{\partial \alpha}{\partial y^o} = -f(y^o) < 0 \tag{5}$$

However, an increase in the acceptance wage increases the wage the participant can expect once employed. In particular,

$$\frac{\partial h}{\partial y^o} = \frac{f(y^o)\int_{y^o}^{\hat{y}} (y - y^o)f(y)dy}{\left[\int_{y^o}^{\hat{y}} f(y)dy\right]^2} \tag{6}$$

$$= \frac{-\left(\frac{\partial \alpha}{\partial y^o}\right)(h - y^o)}{\alpha} > 0$$

The participant chooses his acceptance wage so as to maximize his expected human wealth. Human wealth in this context is the discounted future stream of labor income. The participant will earn a relative wage equal to h in period t if he is employed during the period and if he is still a participant in the labor market in the period t. If he is not a participant, his labor income is zero. But if he is an unemployed participant he may receive unemployment compensation or some other form of income maintenance. Let z denote the ratio of unemployment compensation to the mean wage offered in the market and let w denote the latter. Assume that the probability of being a participant in period t, p_t, is independent of whether or not the participant happens to be employed. Then, if we let q_{t-1} denote the probability of being employed at the beginning of period t and if we assume that the participant can borrow or lend freely at an interest rate r, his expected wealth prior to commencing search is defined by the following equation:[5]

$$W = w \sum_{t=1}^{\infty} \frac{p_t}{(1+r)^t} \cdot [q_{t-1}h + (1 - q_{t-1})z] \tag{7}$$

We have assumed that the event "retiring from the labor force" is stochastic in order to eliminate the analytic complications induced by the alternative assumption of a finite determinant working life. The age of the participant can be introduced into the analysis by assuming the probability of retiring per period, δ, is larger for older workers. Since the expected duration of search will typically be short relative to the expected remaining working life, $1/\delta$, we regard δ as a constant for a given participant. Therefore, p_t equals $(1-\delta)^t$ approximately and

$$\frac{p_t}{(1+r)^t} \cong \left(\frac{1-\delta}{1+r}\right)^t = \frac{1}{(1+\rho)^t} \tag{8}$$

where the new discount rate, ρ, is given by

$$\rho = \frac{r + \delta}{1 - \delta} \tag{9}$$

Therefore, future expected labor income has no present value if the participant is going to retire one period hence.

The probability of being employed at the end of period t is easily ascertained if we choose the period used in computing the expected wealth equal in length to the period required to search a single vacancy. In that case q_t equals unity minus the probability of not finding an acceptable offer in a sample of size t; i.e.,

[5] Implicitly we are assuming that the participant either ignores the possibility of searching *after* becoming employed or considers such search to be prohibitively expensive. Analyses which take the possibility into account include that of Charles C. Holt and Mortensen.

(10) $$q_t = 1 - (1 - \alpha)^t$$

Since $q_t - q_{t-1} = \alpha(1-\alpha)^{t-1}$ is the probability of finding employment during period t, $1/\alpha$ is the expected length of the period spent searching before becoming employed as we asserted earlier.

Substituting for p_t and q_{t-1} in (7) from (8) and (10), respectively, we obtain

$$W = wh \sum_{t=1}^{\infty} \left(\frac{1}{1+\rho}\right)^t - \frac{w(h-z)}{1-\alpha} \sum_{t=1}^{\infty} \left(\frac{1-\alpha}{1+\rho}\right)^t$$

Hence,

(11) $$W = \frac{w}{\rho}\left(\frac{\alpha(y^o, \hat{y})h(y^o, \hat{y}) + \rho z}{\rho + \alpha(y^o, \hat{y})}\right)$$

where the functions $\alpha(y^o, \hat{y})$ and $h(y^o, \hat{y})$ are defined by equations (3) and (4), respectively. By making use of equations (5) and (6), we obtain

$$\frac{\partial W}{\partial y^o} = \frac{wf(y^o)}{\rho[\rho + \alpha(y^o, \hat{y})]^2} \cdot \{\alpha(y^o, \hat{y})[h(y^o, \hat{y}) - y^o] - \rho[y^o - z]\}$$

Because the acceptance wage is that value of y^o which maximizes human wealth, W, the optimal choice of the acceptance wage must satisfy equation (12).[6,7]

(12) $$\alpha(y^o, \hat{y})[h(y^o, \hat{y}) - y^o] = \rho[y^o - z]$$

An economic interpretation of (12) is useful. During the process of search, the participant must decide each time he receives an offer whether to accept it or whether to search another vacancy. Suppose he is offered a relative wage equal to y. If he accepts it, he would receive an income per period equal to y thereafter. If he rejects the offer his income in the next period is the unemployment compensation z. Therefore, the certain cost of searching one more vacancy is $y-z$. If that cost is positive, he will not search another vacancy unless he expects to receive a higher wage than his current offer with some positive probability. Since he expects that the next search will yield a net stream of income equal to $h(y, \hat{y}) - y$ forever with probability $\alpha(y, \hat{y})$, he rejects an offer of y if the expected present value of that stream, $\alpha(y, \hat{y})[h(y, \hat{y}) - y]/\rho$, exceeds the cost of searching another vacancy, $y-z$. In other words, the acceptance wage is that offer which equates the marginal cost of search to the present value of the marginal expected gains from searching.

An interesting interpretation of the acceptance wage arises from the fact that equations (11) and (12) imply

(13) $$W = \frac{wy^o}{\rho}$$

Since expected wealth equals the capitalized acceptance wage, the acceptance wage is equivalent to the concept of permanent labor income.

We must point out that a meaningful solution exists to (12) if, and only if, $\hat{y} \geq z$. To see this more clearly, rewrite (12) as

(14) $$\int_{y^o}^{\hat{y}} (y - y^o)f(y)dy = \rho(y^o - z)$$

by substituting for α and h from equations (3) and (4), respectively. The integral on the left-hand side of (14) has no meaning unless $\hat{y} \geq y^o$. But, there exists no value of y^o satisfying this constraint and equation (14) if $z > \hat{y}$. The economic reason for this

[6] Our formulation of the problem implies that the acceptance wage is independent of the length of time already spent searching for a job. Alternative formulations suggest that the acceptance wage falls as the duration increases. For example, see Holt and Hirschel Kasper.

[7] If a solution to (12) exists, assumption (2a) insures that it is unique and that it maximizes W. The existence problem is examined in the text.

problem is obvious. If $z \geq \hat{y}$, the participant never searches because the best wage he can hope to obtain is less than the certain wage available without searching.

We conclude, then, that the participant will search if, and only if, $\hat{y} > z$. In turn, equation (14) implies

$$\hat{y} > y^o > z \tag{15}$$

if the participant searches. In other words, if the participant does not have perfect knowledge about the nature of particular wage offers, then he will accept some wage less than the highest uncertain possibility and greater than the certain minimum available to him.

It is of interest to note that the proportion of jobs open to a participant, k, is defined as follows:

$$k = Pr(y \leq \hat{y}) = \int_a^{\hat{y}} f(y)dy \tag{16}$$

This proportion and the maximum relative wage attainable are positively related because both are monotone increasing functions of the participant's qualifications. Therefore, the acceptance wage can be expressed as a function of the participant's skill level as measured by the proportion of vacancies open to him, the unemployment compensation or guaranteed income available and the discount rate. In other words,

$$y^o = y^o(k, \rho, z) \tag{17}$$

By completely differentiating (14) and by making use of the fact that $dk = f(\hat{y})d\hat{y}$, we obtain the following results:

$$\begin{aligned} &\text{(a)} \quad \frac{\partial y^o}{\partial k} = \frac{\hat{y} - y^o}{\rho + \alpha} > 0 \\ &\text{(b)} \quad \frac{\partial y^o}{\partial \rho} = \frac{-(y^o - z)}{\rho + \alpha} < 0 \\ &\text{(c)} \quad \frac{\partial y^o}{\partial z} = \frac{\rho}{\rho + \alpha} > 0 \end{aligned} \tag{18}$$

These results are explained by the following facts: An increase in unemployment compensation reduces the cost of search. An increase in the discount rate reduces the present value of future gains to search. The return to searching another vacancy is larger if the proportion of jobs open is larger.

II. *The Duration of Unemployment*

The decision of the unemployed worker concerning his acceptance wage also implies a decision regarding the length of time he can expect to be unemployed in the process of searching for an acceptable job. However, the expected length of the period, $1/\alpha$, depends directly on the proportion of jobs open to the participant as well as the acceptance wage. In particular, equations (3) and (16) imply

$$\alpha = k - \int_a^{y^o} f(y)dy \tag{19}$$

Because the optimal acceptance wage is a function of the discount rate, the level of unemployment compensation and the proportion of jobs open to the *participant*, the expected duration of search unemployment is also a function of these same variables. In particular, an increase, *ceteris paribus*, in the discount rate reduces the duration, while an increase, other things constant, in unemployment compensation lengthens it because these changes induce, respectively, a reduction and an increase in the optimal acceptance wage. An increase in the proportion of jobs open to the participant has two effects on the expected duration. The direct effect, i.e., the effect when the acceptance wage is held constant, lowers the duration. However, equation (18a) implies that an increase in the proportion induces an increase in the acceptance wage because such a change reflects the fact that the participant's maximum attainable wage

has increased. Consequently, the indirect effect tends to offset the reduction on duration attributable to the direct effect.

The ambiguity of the response in α with respect to an increase in the proportion of jobs open needs some further analysis. Equation (18a) and (19) imply

$$\begin{aligned}\frac{\partial\alpha}{\partial k} &= 1 - f(y^o)\frac{\partial y^o}{\partial k} \\ &= 1 - \frac{f(y^o)(\hat{y} - y^o)}{\rho + \alpha}\end{aligned} \tag{20}$$

Since $\hat{y} = y^o$ when $\hat{y} = z$, (20) implies that $\partial\alpha/\partial k = 1$ when the qualifications of the participant are such that his best offer equals the unemployment compensation available. The continuity of α with respect to k implies, then, that an increase in the proportion of jobs open decreases the duration of search for those participants who are qualified for a sufficiently small proportion of the existing vacancies.

One further result is important because it suggests that the sign of the response in the duration of employment with respect to an increase in skill, as measured by the proportion of jobs open to a given participant, may depend on the nature of the upper tail of the wage distribution. The probability of finding an acceptable job per period, α, is a monotone increasing function of the proportion of jobs open to a participant if that participant's maximum attainable wage, $\hat{y}$, and the distribution of relative wages is such that the following condition is satisfied:

$$f(y) + yf'(y) \geq 0 \quad \text{for all } y^o \leq y \leq \hat{y} \tag{21}$$

To prove the assertion, we use the fact that (21) implies $yf(y) > y^o f(y^o)$ for all $y^o \leq y \leq \hat{y}$. It does so because $f(y) + yf'(y)$ is the derivative of $yf(y)$. Then, because equations (14) and (3) imply

$$y^o(\rho + \alpha) - \rho z = \int_{y^o}^{\hat{y}} yf(y)dy,$$

condition (21) implies

$$y^o(\rho + \alpha) \geq \int_{y^o}^{\hat{y}} yf(y)dy > y^o f(y^o)(\hat{y} - y^o) \quad \text{for all } y^o \leq y \leq \hat{y}$$

Given this inequality, the assertion follows from equation (20). Note that (21) is satisfied given a unimodal distribution if $\hat{y}$ is less than or equal to the mode of the distribution since $f'(y) > 0$ for all $y^o \leq y \leq \hat{y}$. Of course, (21) is satisfied no matter what the value of y if the distribution is uniform since $f'(y) = 0$ for all y.

We have established that a participant searches, i.e., $\alpha > 0$ if, and only if, the maximum wage attainable is greater than any unemployment compensation. Since we have shown that the proportion of jobs open to a given participant depends only on the maximum wage attainable (see equation (16)), the condition can as well be expressed as $k > k_0$ where k_0 is the value k when $\hat{y} = z$. In other words, the proportion of jobs open to the participant must exceed the proportion of vacancies which offer a wage equal to or less than the unemployment compensation available. If we define k_1 as the proportion of vacancies which offer a wage equal to or less than the mode of the wage distribution, then a marginal increase in the proportion of vacancies for which the participant qualifies results in a decrease in the expected duration of search if it is less than or equal to the k_1 and greater than or equal to k_0. These results together with those already discussed are summarized below.

(22)

(a) $\alpha = \alpha(k, \rho, z)$;

$$\alpha = \begin{cases} 0 & \text{if } k \leq k_o \\ >0 & \text{if } k > k_o \end{cases}$$

(b) $\dfrac{\partial\alpha}{\partial k} > 0 \quad \text{if } k_o \leq k \leq k_1$

(c) $\dfrac{\partial\alpha}{\partial\rho} = -f(y^o)\dfrac{\partial y^o}{\partial\rho}$

$$\text{(c)} \qquad = \frac{f(y^o)(y^* - z)}{\rho + \alpha} \geq 0 \quad \text{as } k \geq k_*$$

$$\text{(d)} \quad \frac{\partial \alpha}{\partial z} = -f(y^o)\frac{\partial y^o}{\partial z} = -\frac{f(y^o)\rho}{\rho + \alpha} < 0$$

The aggregate implications of our model of job search behavior are of interest at this point. In aggregating we cannot make the convenient assumption that all unemployed participants are identical. If all were identical each would be qualified for all openings, firms would offer the same wage for all openings and there would be no reason to search. However, let us abstract somewhat by ignoring differences in the discount rate and in the unemployment compensation available to different participants. Our presumption is that participants search for periods of different lengths primarily because of the differences among them with respect to skill.

Assume that skills are distributed such that the distribution of k over the population of participants can be described by a continuous density function which we denote as $j(k)$. Then the average value of α, which we denote as $\bar{\alpha}$, is given by

$$\text{(23)} \qquad \bar{\alpha} = \int_0^1 \alpha(k, \rho, z) j(k) dk$$

In general, then, $\bar{\alpha}$ depends on the nature of the distribution of skills, the discount rate and the level of unemployment compensation. The interpretation of $\bar{\alpha}$ is obvious. It is the expected proportion of the stock of unemployed participants who find employment per period.

The rate of change in the stock of unemployed participants equals the flow of new entrants into the labor force less the flow of unemployed participants who either retire or become employed. If the retirement rate equals a constant, δ, then the latter flow equals $(\delta+\bar{\alpha})U$ where U denotes the number of unemployed participants. If a constant proportion of the working age population participates in the labor force and if the size of that population grows at a gross proportional rate, η, then the flow of new entrants equal ηL where L is the total number of employed and unemployed participants.[8] If the retirement rate is the same whether participants are employed or not, the total labor force grows at a net proportional rate equal to $\eta-\delta$. Thus, the number unemployed remains constant relative to the total number in the labor force when

$$(\eta - \delta)U = \eta L - (\bar{\alpha} + \delta)U$$

In other words, the equilibrium unemployment ratio, u, that value of U/L which satisfies the equation above, equals the proportion of the labor force which enters per period divided by the sum of that proportion and the proportion of unemployed which finds employment during the period, i.e.,

$$\text{(24)} \qquad u = \frac{\eta}{\eta + \bar{\alpha}}$$

We conclude, then, that some proportion of the labor force will be unemployed in equilibrium if the flow of new entrants is positive and if some time is required to find an acceptable job.

Within the context of the model presented here, two conditions are necessary in order that the average time spent searching $1/\bar{\alpha}$, be positive. First, the participants must differ in terms of their qualifications for employment; otherwise there would be no sustainable reason for employers to maintain wage differentials. Second, some degree of imperfect information must exist in the sense we have defined it. If the participant knew the particulars concerning each opening, he would immediately

[8] This assumption is equivalent to assuming that the supply curve is perfectly inelastic with respect to the wage rate.

accept that job offering the highest wage consistent with his skill.

Of course, equation (24) implies that any change which induces a decrease, *ceteris paribus*, in the average duration of search will decrease the equilibrium unemployment ratio. Since we have already discussed the relationship between the duration of unemployment and its determining factors for a particular participant, there is no need to discuss the relationship among these factors and the equilibrium unemployment ratio. However, one point is worth mentioning. Since the response of the duration with respect to an increase in the proportion of jobs cannot be signed in general, our theory does not imply that any policy designed to increase this proportion, such as a retraining program, will decrease the equilibrium unemployment ratio. However, if such a program were restricted to participants whose maximum attainable wage is less than or equal to the mode of the wage distribution, the theory does suggest that the unemployment ratio will fall as a consequence of the program.

III. *The Phillips Curve*

In the preceding section an equilibrium unemployment ratio was derived with virtually no explicit consideration of the demand side of the market. Here we attempt to bring the demand side of the labor market into the analysis. The consequence of doing so is a complete dynamic model of wage and unemployment adjustment. The unemployment ratio derived in the preceding analysis is the equilibrium state of that model. In addition, the model provides an explanation of the Phillips curve. It is a set of observations consistent with adjustment to the equilibrium unemployment ratio.

The primary objective of the section is to explain the posited relationship between relative wage offers and the degree of qualification required, and to provide a theory of how that relationship shifts as general economic conditions change. The relationship is essentially a description of the structure of wage rates. Our contention is that equation (1) describes only "long run" or equilibrium structure. If in the "short run" there is a general increase in demand for the products of firms, then each firm lowers the skill requirements attached to its vacancies and attempts to raise its wage offer relative to all others in the market in order to attract labor at a more rapid rate. The consequence of this joint strategy is that wage inflation is systematically associated with general lowering of minimum skill requirements. Of course, a general lowering of skill required implies a general decrease in the duration of search unemployment. A Phillips curve results from the operation of the process.

Clearly, the description of the search process presented in Section I implies that within each period a limited number of unemployed workers search any one firm or establishment. In other words, the supply of labor available to a particular establishment in a given period is limited to the stock of its own employees plus the flow of those unemployed participants who happen to inquire about job vacancies within that period. Since some proportion of those currently employed will retire or quit for other reasons, some of those searching are hired if the managers of the firm do not wish to decrease the stock of employees by the number who leave. Since the group of unemployed workers who search the establishment within the period do not all have the same qualification or the same acceptance wage, the proportion hired is regulated by the firm's choices concerning its own relative wage and the minimum level of skill it requires.

In particular, if the chance that an unemployed participant searches the ith es-

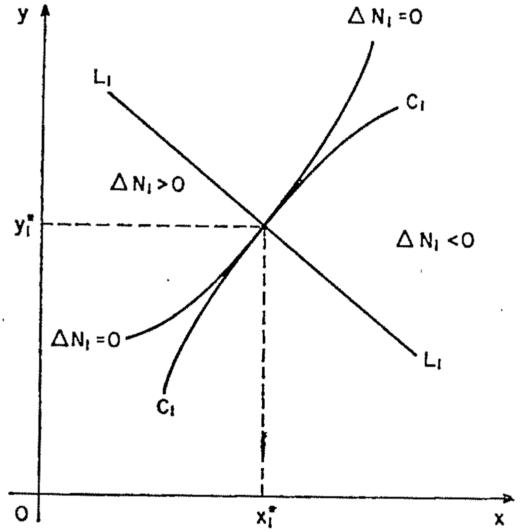


FIGURE 2

tablishment is proportional to the relative size of the establishment, as measured by employment, then the number searching that establishment per period is proportional to the employment of the ith establishment, N_i. Let $\theta(y_i, x_i)N_i$ denote the number of these hired.[9] Since more are hired the higher the establishment's relative wage offer, y_i, and the lower the minimum level of skill required, x_i, the partial derivatives of θ are positive and negative, respectively. Finally, if the number of existing employees who retire equals δN_i, then the proportional rate of change in the establishment's labor force is given by

$$(25) \quad \Delta N_i/N_i = \theta(y_i, x_i) - \delta \quad \theta_1 > 0, \ \theta_2 < 0$$

We assume throughout the remainder of the analysis that the form of the function θ is the same for all firms. Thus, any two establishments which offer the same wage and require the same level of skill hire at the same net proportional rate.

The managers of the firm face two interrelated choice problems. On the one hand, they must take a decision on the number to hire and, on the other, they must determine the optimal combination of offer and minimum skill requirement to use in order to obtain that number. These decisions, of course, must be taken each period.

The problem of choosing the optimal combination (y_i, x_i) is analytically analogous to the problem of choosing input combination in the theory of production. Equation (25) and the function θ imply the existence of sets of points in the yx plane all of which yield the same proportional rate of change in the firm's employment given δ. If we assume that θ is continuous with continuous first partial derivatives, these sets can be represented by a family of nonintersecting curves all of which have positive slopes. In Figure 2 we have drawn a representative curve which, for expositional purposes, we interpret to represent the set of points corresponding to no change in the stock of employees. Thus, all curves above and to the left of that curve correspond to positive net rates of employee accumulation while those below and to the right represent various rates of deaccumulation.

Assume that the firm in question wishes to accumulate employees at some given rate. How does it choose a particular combination from among the set of possibilities which yield that rate? Presumably that one which minimizes current outlay is selected. Clearly costs are involved. A higher wage offer implies a larger wage bill. Given the offer, a reduction in the minimum skill level accepted results in additional cost either in the sense that productivity falls or that expenditure is

[9] At this point the reader is probably aware of the fact that the term "vacancy" is used rather loosely in Section I. Our establishment hires all applicants who qualify and who are willing to accept the wage offered. The unit sampled is not the job opening; it is the establishment or firm. However, this formulation is consistent with Section I if the distribution of wage offers is constructed by weighting the ith firm's offer by the chance of sampling it. Our assumption is that the establishment's share of total employment is the appropriate weight.

required to train the new employees to the same skill as existing employees. There exist, then, a family of positively sloped iso-cost curves in the yx plane. In addition, current outlay is generally higher for points above and to the left of any point in the plane.

Suppose that no change in the stock of employees is desired. The optimal combination is that point on the $\Delta N_i=0$ curve at which that curve is tangent to the lowest possible iso-cost curve. If such a point exists in the interior of the yx positive quadrant, the $\Delta N_i=0$ curve and the iso-cost curves must have, of course, certain relative curvature properties. In Figure 2 we have drawn the iso-cost curve, C_iC_i, corresponding to the minimum cost of maintaining a constant labor force with these properties. Thus, the optimal equilibrium combination for our establishment is (y_i^*, x_i^*). By solving the cost minimization problem for all possible rates of accumulation and deaccumulation, we can generate a locus of points of tangency which provides us with the combinations of y and x which minimize costs under all circumstances. This locus is likely to have a negative slope as illustrated in Figure 2 by the curve labeled L_iL_i. The negative slope simply indicates that the firm raises its relative wage offer and lowers the minimum skill requirements when it desires to accumulate employees at a more rapid rate.

How can we square our assertion that a given firm will lower the qualifications required when it raises the wage it offers with the assumption made earlier (equation (1)) that the relationship between relative offers available and the level of minimum qualifications as viewed by unemployed participants has a positive slope? This apparent paradox is solved by realizing that different firms produce different products and that the technologies used to produce these different products generally require labor input with different skills and by realizing that all establishments ***in equilibrium*** wish to maintain a constant labor force. In other words, in equilibrium each establishment will choose some point on the $\Delta N_i=0$ curve but not the same point. Those establishments which have technologies that require more highly skilled labor choose points higher on the curve in order to avoid prohibitive training costs or low productivity. Although some establishments may choose points above and others below the curve during some arbitrarily chosen period, the general relationship between all relative wage offers and their corresponding minimum skill requirements will exhibit the positive slope assumed. In other words, if the net changes in employment across establishments are uncorrelated, a regression line would yield the set of points in the yx plane consistent with no change in employment by any firm.

An important implication follows from this analysis. If the changes in employment across firms are correlated in a given period, i.e., if all or most firms desire to accumulate or deaccumulate employees at similar rates, the aggregate relationship between the relative wage offers and minimum qualifications shifts in a more or less consistent manner. Our next task is to derive a proxy which signals shifts occurring for this reason. To do so we must consider the second aspect of each firm's decision problem.

The problem of choosing the optimal rate of change in the stock of employees is essentially one in the theory of investment. Because the firm must offer a higher wage and spend more on training, current outlay is higher the more rapid the rate of accumulation. The firm incurs these costs up to the point where the investment made to hire the marginal new employee equals the future discounted stream of returns imputable to that new

employee.[10] This condition implies, in terms of the familiar stock adjustment model, that the firm desires in the long run to acquire a stock of employees of a particular size. The equilibrium stock is that which equates the value of the marginal product of labor to the equilibrium wage rate plus the cost of maintaining the labor force equal to the equilibrium level. In each period, the firm adjusts its labor force only partially to the equilibrium level.

Since the value of the marginal product of labor is proportional to the price of the firm's product, P_i, and since its equilibrium wage is proportional to the average wage, w, any particular firm will accumulate employees at a more rapid rate given an increase, *ceteris paribus*, in the ratio P_i/w. It does so because the equilibrium stock of employment increases. Of course, given the value of this ratio, the rate of accumulation falls as the equilibrium employment stock is approached. In other words, the rate of change in the stock diminishes as the size of the existing labor force increases, *ceteris paribus*.

Suppose that the prices of the products produced by all firms in a given labor market were to rise. The result would be a general attempt by all firms to accumulate employees at a more rapid rate. In other words, each firm would lower the minimum skill level that it required and attempt to raise its own wage offer relative to the average prevailing in the market. Of course, not all can raise their money wage relative to the average money wage offer. The result of attempting to do so is a general increase in the wage level. Thus, the rate of wage inflation is the appropriate indicator of a shift in the relationship between relative wage offers and skill requirements.

Actually the difference between the rate of money wage inflation and the expected rate of product price inflation is the appropriate proxy. To see this point, suppose that each employer expects the average wage to be 3 percent higher this period than last, and assume that the price of each product has increased sufficiently enough to justify a 5 percent increase in the wage each employer offers relative to the average. The consequence of each attempting to offer a wage 8 percent higher than the previous one is an 8 percent increase in the average wage; an increase which is 5 percent higher than that which each had anticipated. To clinch the argument, suppose instead that each firm expects the price of its product to increase by 3 percent and the average wage to rise by 3 percent in the current period. Assume that all firms currently have their equilibrium labor force. Since each firm expects its product price divided by the average wage to remain constant, none attempt to accumulate employees. As a result, each offers a 3 percent increase in its money wage relative to that offered in the previous period in order to maintain its relative wage offer. Thereby, expectations are confirmed. Clearly no shift has occurred in the relationship across firms between the relative wage offer and the minimum level of skill needed. We conclude then that the disequilibrium relationship between relative wage offers and skills can be written as

$$(26) \quad y = y(x, g - p^e); \quad y_1 > 0,\ y_2 > 0,$$

where

$$(27) \quad \begin{aligned} &(a) \quad g = \frac{\Delta w}{w} \\ &(b) \quad p = \frac{\Delta P}{P} \end{aligned}$$

[10] A complete analysis is presented in Mortensen. The model used is analytically equivalent to cost of adjustment models of investment behavior. See, for example, Robert Eisner and Robert Strotz, John Gould, Robert Lucas, and Arthur Treadway. Costs of adjusting the level of employment have also been used by Frank Brechling and Robert Solow to rationalize the behavior of employment over time.

where p^e is the rate of inflation anticipated by employers and P is an index of product prices.

Clearly, we have not only solved our original problem. In the process we have described the mechanism by which wages are bid up and down in our market. In particular, when all or most employers simultaneously experience an increase in the prices of their products relative to the average wage or more accurately when the equilibrium desired levels of employment increase generally relative to the existing levels for any reason, the wage level is bid up relative to that expected as a consequence of each attempting to attract more employees. Conversely, the rate of wage inflation falls relative to that expected as employment approaches equilibrium. Mathematically, we can express this relationship as

$$\frac{\Delta w}{w} = g = p^e + \theta(P/w, N/L) \tag{28}$$

$$\theta_1 > 0,\ \theta_2 < 0,$$[11]

where N is total employment.

The reason periods of money wage inflation should also be periods of falling unemployment is now clear. If a given period is one of general excess demand for labor in the sense that all or most firms wish to accumulate a force larger than their present one, each firm attempts to attract new employees by lowering its skill requirements and by attempting to raise its wage offer relative to all others. The effect of the latter strategy is money wage inflation. The effect of the former is an increase in the proportion of jobs open to every unemployed participant. Even if the increase in the proportion were perceived, the duration of search by each unemployed participant is likely to fall.

Our assumption is that the participant cannot predict the periods of general excess demand or supply. In other words, he uses the long-run relationship between offers and qualifications as a basis for his decision. Therefore, in his view the proportion of jobs open to him depends only on his own qualifications.

Since the actual maximum relative wage available to an unemployed participant whose skill level is x^o is $y(x^o, g-g^e)$, the actual probability of finding a job per period is given by

$$\begin{aligned} \beta &= \int_{y^o}^{y(x^o,g-p^e)} f(y)dy \\ &= \int_{y^o}^{y(x^o,0)} f(y)dy + \int_{y(x^o,0)}^{y(x^o,g-p^e)} f(y)dy \\ &= \alpha(k, \rho, z) + \int_{y(x^o,0)}^{y(x^o,g-p^e)} f(y)dy \end{aligned} \tag{29}$$

If inflation is not fully anticipated; i.e., if $g > p^e$, the second term is positive. The term represents the additional probability resulting from the fact that skill requirements have been lowered.

If the difference between the actual and anticipated rates of inflation is small, the following equation holds as an approximation:

$$\beta = \alpha + \alpha_o(g - p^e) \tag{30}$$

where

$$\alpha_o = y_2(x^o, 0)f(y(x^o, 0)) > 0 \tag{31}$$

Since α_o depends on the qualifications of the participant as well as α, the expected proportion of all unemployed workers who find employment per period is

$$\begin{aligned} \bar{\beta} &= \int_0^1 \beta j(k)dk \\ &= \int_0^1 \alpha j(k)dk + (g - p^e)\int_0^1 \alpha_o(k)dk \\ &= \bar{\alpha} + (g - p^e)\bar{\alpha}_o \end{aligned} \tag{32}$$

Therefore, the change in unemployment is given by

$$\Delta U = \eta L - (\bar{\beta} + \delta)U \tag{33}$$

[11] For a complete derivation see Mortensen.

where ηL is the flow of new entrants and δU is the flow of retirements from the stock of unemployed. Let $u=U/L$ be the unemployment ratio. Because

$$\Delta u/u = \Delta U/U - \Delta L/L$$

approximately and because the net proportional change in participation is $\eta-\delta$, equations (32) and (33) imply

$$(34) \quad \Delta u/u = \eta/u - \eta - \bar{\alpha} - \bar{\alpha}_o(g - p^e)$$

An increase, *ceteris paribus*, in the rate of money wage inflation decreases the proportional rate of change in the unemployment ratio.

A Phillips curve can be obtained simply by rewriting equation (34) as

$$g = p^e + \frac{1}{\bar{\alpha}_o}\,[\eta/u - \eta - \bar{\alpha} - \Delta u/u]$$

Given the anticipated rate of inflation and the proportional rate of change in the unemployment ratio, the rate of money wage inflation and the unemployment ratio are inversely related. This result follows from the fact that an increase in either g or u increases the flow of unemployed participants who find acceptable employment per period.

IV. *Summary*

The Phillips curve, equation (34), is one of three which form a simultaneous equations model of the dynamic behavior of the labor market. It describes the manner in which the unemployment ratio responds over time to conditions in the market. Equation (28) describes the dynamics of money wage adjustment. If we assume that the rate of product price inflation is exogenous, a third equation specifying how the expected rate responds to changes in the actual rate is needed to complete the model.

Clearly the rate of money wage inflation and the unemployment ratio are endogenous to this model. Therefore, different observations on these variables are generated by changes which are exogenous to the model. Rather obvious sources of disturbance are changes in demand for the products of the firms employing labor in this market. It can be shown that fluctuations in the rate of product price inflation, given this model, can produce data consistent with the original observations of A. W. Phillips.[12] In addition, the counterclockwise loops observed by Phillips and discussed by Richard Lipsey are implied by the model. The only necessary condition is that the expected rate of inflation not adjust instantaneously to changes in the actual rate. Essentially the argument is that fluctuation in product demand will shift about the upward sloping relationship between the rate of wage inflation and the unemployment ratio implied by equation (28). Therefore, regressing the unemployment ratio on the rate of money wage inflation identifies the inverse relationship implied by equation (34).

Although the model is consistent with the existence of a Phillips curve in the sense of a set of observed values of the rate of money wage inflation and the unemployment ratio which are inversely related, the model also implies that the phenomenon is a transitory one. The model possesses a unique equilibrium state. In equilibrium the following relationships hold:

$$(35) \quad \begin{aligned} &\text{(a)} \quad p = g = p^e \\ &\text{(b)} \quad \theta(P/w, 1 - u) = 0 \\ &\text{(c)} \quad u = \frac{\eta}{\eta + \bar{\alpha}} \end{aligned}$$

In equilibrium real wages are on average constant so that the rate of product price inflation equals the proportional rate at which the average money wage inflates.

[12] See Mortensen.

In addition, inflation is anticipated. In other words, the real wage levels have adjusted so that the firms on average are just replacing those who retire with new employees. They do so by maintaining their relative wage offers and the skills they require at their equilibrium values. Since these are at their equilibrium values, the expected durations of search unemployment are equal to those planned by the unemployed participants. Thus, new entrants into the labor market spend an average $1/\bar{\alpha}$ periods before they locate an acceptable job.

Because new entrants continuously flow into the market and because they view some time spent searching for a job as a worthwhile investment in a world of imperfect information, there are always new entrants searching for a job. The determinants of the equilibrium number of unemployed as a proportion of the total labor force have already been discussed. The important implication here is that the proportion is independent of the rate of inflation. In other words, the model is consistent with the assertions made by Phelps (1968a) and Milton Friedman that there exists an "equilibrium," in Phelp's language, or "natural," as Friedman has labeled it, unemployment ratio which is independent of the inflation rate.

REFERENCES

A. Alchian, "Information Costs, Pricing and Resource Unemployment" in E. Phelps et al., *The Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

F. P. R. Brechling, "The Relationship Between Output and Employment in British Manufacturing Industries," *Rev. Econ. Stud.*, July 1965, *91*, 187–216.

P. H. Cagan, "The Monetary Dynamics of Hyperinflation" in M. Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956.

R. Eisner and R. H. Strotz, "Determinants of Business Investment" in Commission on Money and Credit, *Impact of Monetary Policy*, Englewood Cliffs 1963.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

J. P. Gould, "Adjustment Cost in the Theory of Investment of the Firm," *Rev. Econ. Stud.*, Jan. 1968, *35*, 47–55.

C. C. Holt, "Job Search, Phillips' Wage Relation, and Union Influence" in E. Phelps et. al., *The Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

H. Kasper, "Asking Price of Labor and the Duration of Unemployment," *Rev. Econ. Statist.*, May 1967, *49*, 165–72.

R. G. Lipsey, "The Relation Between Unemployment and the Rate of Change in the Money Wage Rates in the United Kingdom 1862–1957: A Further Analysis," *Economica*, Feb. 1960, *27*, 1–31.

R. E. Lucas, Jr., "Optimal Investment Policy and the Flexible Accelerator," *Int. Econ. Rev.*, Feb. 1967, *8*, 78–85.

——— and L. A. Rapping, (1969a) "Price Expectations and the Phillips Curve," *Amer. Econ. Rev.*, June 1969, *59*, 342–50.

——— and ———, (1969b) "Real Wages, Employment and Inflation," *J. Polit. Econ.*, Sept./Oct. 1969, *77*, 721–54.

D. T. Mortensen, "A Theory of Wage and Employment Dynamics" in E. Phelps, et. al., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

G. L. Perry, "The Determinants of Wage Rate Changes," *Rev. Econ. Stud.*, Oct. 1964, *31*, 287–308.

———, *Unemployment, Money Wage Rates, and Inflation*, Cambridge, Mass. 1966.

E. S. Phelps, (1968a) "Money Wage Dynamics and Labor Market Equilibrium," *J. Polit. Econ.*, Aug. 1968, *76*, 687–711.

———, (1968b) "Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time," *Economica*, Aug. 1968, 254–81.

———, "The New Microeconomics in Inflation and Employment Theory," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 147–60.

———, et al., *Microeconomic Foundations of Employment and Inflation Theory,* New York 1970.

A. W. Phillips, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957," *Economica,* Nov. 1958, *25,* 283–99.

R. Solow, "Short-Run Adjustment of Employment to Output" in J. N. Wolfe, ed., *Value, Capital and Growth,* Edinburgh 1968.

G. J. Stigler, "Information in the Labor Market," *J. Polit. Econ.,* Oct. 1962, *70,* 94–105.

———, "The Economics of Information," *J. Polit. Econ.,* June 1961, *69,* 213–225.

A. B. Treadway, "On Rational Entrepreneurial Behavior and the Demand for Investment," *Rev. Econ. Stud.,* Apr. 1969, *36,* 227–39.

Value-Added and Factor Productivity in Soviet Industry

By JUDITH THORNTON*

This study develops and uses new estimates of value-added in Soviet industry in the postwar period. These value-added estimates represent an alternative to available series on the growth of Soviet industrial output that have been constructed from gross output data for industry.

The Soviet official measure of industrial growth is summarized in an index of gross value of output at constant prices. However, the usefulness of the official index is reduced by the lack of information about computation procedures and by its inconsistency with other published information on the value of output.

Alternative measures of Soviet industrial output have been presented in a number of Western studies. These studies present gross output series built up from value weighted samples of individual commodities measured in physical units. (See studies by Rush Greenslade, Norman Kaplan and Richard Moorsteen, James Noren, Warren Nutter, Raymond Powell.)[1] Commodities are aggregated within industries by product prices and among industries by hypothetical value-added where the latter is constructed from industrial wages bill plus some assumed rate of return applied to an estimate of capital in industry.

This study provides an alternative method of estimating industrial value-added at current and constant prices based on officially published data from the Soviet cost accounts. The value-added series are broken down to provide estimates of the shares of labor and capital in value-added. This breakdown indicates that the share of capital in Soviet value-added is larger than has been estimated in earlier studies and that it is growing during the post war period.

Then, input data are combined with the new value-added data to estimate the implied contributions of inputs and of total factor productivity in accounting for the growth of industrial output.

Data supporting the estimates presented here are available in an appendix which can be requested from the author. References to tables in this appendix are identified by an A prefix.

I. *Industrial Gross Output at Current and Constant Prices*

In 1967, the Soviets for the first time published a partial series on the value of industrial output at current enterprise prices. These official data provided the first cross-check of the cost account estimates of the value of industrial output that I had prepared earlier (see Thornton, 1965). They showed that the cost account procedures yielded close estimates—within 3 percent—of the officially calculated value of industrial output at current enterprise prices.

The procedures used in deriving the cost account estimates of the value of industrial output have been described in my earlier article. They involve applying an estimate of the ruble value of wages bill

* University of Washington, Seattle.

[1] The Powell series is estimated somewhat differently, from the industrial component of national income by final use.

in industry to the officially published percentage structure of cost in industry to get an estimate of the total cost of industrial output. Then, total cost and profit in industry are summed to yield the value of industrial output at current enterprise prices (net of turnover taxes.)

According to these series, the value of industrial gross output at current prices grew from 97 to 286 billion rubles between 1955 and 1967, or at an average annual rate of 9.4 percent.

The period of 1955 to 1966 produced relative price stability for the industrial sector as a whole. The Soviet official price index shows a modest rise in the average level of prices until 1966 and a sharp increase in 1967. The price index with a 1955 base stands at 104 in 1966 and at 113 in 1967 (Table A-2).

Deflating the value of output at current prices by the Soviet official price index yields estimates of the value of industrial gross output at constant prices (Table A-1). For the period 1955–67, gross output at constant prices shows an average annual rate of growth of 8.3 percent. This constant price output series agrees surprisingly closely with the high growth variant of the estimates of Soviet industrial growth made by James Noren for the Joint Economic Committee (p. 280). Using his high growth variant for the period 1955–59, the Noren series yields an average annual growth rate for the whole period of 8.4 percent. The Joint Economic Committee index of industrial production approximates a value-added weighted production index like that of the U.S. Federal Reserve Board. According to Evsey Domar, such an index proves a better approximation to gross output than to value-added, which turns out to be the case here.

However, the deflated value of industrial gross output series derived here grows more slowly than the published Soviet official series on the value of industrial output at constant prices of July, 1955 which shows an average growth rate of gross output of 9.5 percent for the period. The quantity weights used in deriving the Soviet official price index are not clearly described, but recent Soviet sources indicate that the official price indices have Paasche, or late period quantity weights (see T. V. Riabushkin, pp. 78–86), (S. G. Stoliarov, pp. 86–113). In that case, the constant-price output series derived here would appear to be a Laspeyres index, so the difference between it and the Soviet official constant-price index cannot be explained by simple index number bias alone.[2]

II. *Industrial Value-Added at Current and at Constant Prices*

Using the cost account structure of output described above, I can separate the gross value of output into value-added and purchased inputs and partition value-added, in turn, into *wages bill, profit, amortization,* and *other cost.* These estimates are presented in Table 1.

The *other cost* category is small—3 percent of the total cost—but it presents real problems. Some of these miscellaneous costs—interest on bank credit, government fines and penalties, rental payments—should be allocated to capital costs and rents; some—bonuses, payments of travel cost of procurement agents, payment of repair costs to subsidiary organizations—are mainly payments to labor. To further complicate matters, some of these costs belong in value-added and some are purchased from outside organizations. In

[2] The coverage of the Soviet official price index is limited—278 products until 1961 and 961 products, or 56 percent of the value of output, after 1961. To the extent that the official price index fails to measure the price rise in new or uncovered products, the growth of the constant-price output series will be biased upwards. An excellent discussion of this problem is presented in R. V. Greenslade.

TABLE 1—INDUSTRIAL VALUE-ADDED—COST ACCOUNT ESTIMATES
(Billion Current Rubles)

	1955	1956	1957	1958	1959	1960	1961
Value of industrial output at enterprise prices[a]	(96.864)	(104.245)	(112.444)	(128.201)	(139.344)	157.4	172.6
Profit in industry[b]	5.949	6.6	8.2	10.677	12.515	14.017	15.297
Cost of industrial output[c]	90.915	97.645	104.244	117.524	126.829	143.383	157.303
Purchased inputs	66.186	70.988	75.681	86.498	93.853	106.103	116.719
(Percentage of cost)	(72.8)	(72.7)	(72.6)	(73.6)	(74.0)	(74.0)	(74.2)
Amortization	3.000	3.320	3.544	3.996	4.439	5.018	5.663
(Percentage of cost)	(3.3)	(3.4)	(3.4)	(3.4)	(3.5)	(3.5)	(3.6)
Wages and social insurance	19.001	20.408	21.787	23.270	24.478	27.673	29.730
(Percentage of cost)	(20.9)	(20.9)	(20.9)	(19.8)	(19.3)	(19.3)	(18.9)
Other cost	2.727	2.929	3.232	3.851	4.058	4.588	5.191
(Percentage of cost)	(3.0)	(3.0)	(3.1)	(3.2)	(3.2)	(3.2)	(3.3)
Value-added excluding other cost	27.951	30.328	33.531	37.943	41.433	46.709	50.690
Share of labor in value-added	.68	.67	.65	.61	.59	.59	.59
Value-added from[d] national income				38.2	45.4	49.9	54.3
Share of labor in value-added				.61	.54	.55	.55

	1962	1963	1964	1965	1966	1967
Value of industrial output at enterprise prices[a]	188.4	201.0	212.4	229.4	248.3	285.9
Profit in industry[b]	18.586	19.600	21.912	22.548	25.088	34.899
Cost of industrial output[c]	169.814	181.400	190.488	206.852	223.2	251.0
Purchase inputs	126.511	133.873	139.818	152.450	164.498	187.748
(Percentage of cost)	(74.5)	(73.8)	(73.4)	(73.7)	(73.7)	(74.8)
Amortization	6.283	8.526	9.334	10.343	11.606	12.299
(Percentage of cost)	(3.7)	(4.7)	(4.9)	(5.0)	(5.2)	(4.9)
Wages and social insurance	31.585	33.196	35.050	37.233	39.953	42.419
(Percentage of cost)	(18.6)	(18.3)	(18.4)	(18.0)	(17.9)	(16.9)
Other cost	5.434	5.805	6.286	6.826	7.142	8.534
(Percentage of cost)	(3.2)	(3.2)	(3.3)	(3.3)	(3.2)	(3.4)
Value-added excluding other cost	56.455	61.322	66.296	70.124	76.660	89.617
Share of labor in value-added	.56	.54	.53	.53	.52	.47
Value-added from[d] national income	59.7	66.0	70.0	72.1	76.9	88.0
Share of labor in value-added	.53	.50	.50	.52	.52	.48

[a] Data for 1960–67 from *SSSR v tsifrakh v 1967 godu* (Moscow: Statistika, 1968), p. 50; 1968 from *Narodnoe Khoziaistvo*, 1968, p. 183. Estimates of gross output for the years 1955–59 are the author's estimates based on cost account data. Soviet official data on employment, average wage, and social insurance rates are used to derive an estimate of wages bill in industry. The ruble value of wages bill is compared with the Soviet official percentage structure of cost to yield the cost of industrial output. And cost plus Soviet official profit yields gross output at enterprise prices. The cost account series yield estimates approximately 3 percent lower than the Soviet official series for the period 1960–67. Employment and labor data used in the cost account estimates are based on *Trud v SSSR* (Moscow: Statistika, 1968) pp. 24–29 and 81–89; 136–45.

[b] *Narodnoe Khoziaistvo* 1965, p. 757; *Promyshlennost'* 1964, pp. 100–1; A. G. Zverev *Problemy' tsenoobrazovaniia i finansy* (Moscow: Nauka, 1966) *SSSR v tsifrakh v 1967 godu*, p. 28.

[c] Cost account structures from annual *Narodnoe Khoziaistvo* 1956, p. 93; 1958, p. 171; 1959, p. 161; 1960, p. 240; 1961, p. 195; 1962, p. 143; 1963, p. 135; 1964, p. 153; 1965, p. 165; *Strana Sovetov za 50 let*, p. 64.

[d] *Narodnoe Khoziaistvo* 1965, p. 591; 1964, p. 577; 1963, p. 502; 1962, p. 482; 1960, p. 153; 1967, p. 672. *SSSR v tsifrakh v 1966 godu* (Moscow: Statistika, 1967) p. 33; *Narodnoe Khoziaistvo* 1965, p. 757; 1964, p. 747; 1963, p. 637; 1962, p. 635; 1961, p. 755. Turnover tax in industry is estimated to be 98.6 percent of all turnover tax, from M. Eidel'man "Metodicheskie voprosy analiza balansa narodnogo khoziaistvo."

TABLE 2—VALUE-ADDED IN INDUSTRY
(Billion Current Rubles)

Other Western Estimates			Thornton Estimates: Cost Account	Thornton Estimates: National Income
1955 (Nutter)				
Value-Added	25.8		28.0	
Wages Bill	15.0		19.0	
Returns to Capital	10.8		9.0	
1959 (Cohn)				
Value-Added	41.6		41.4	45.4
Wages Bill	28.6		24.5	24.5
Returns to Capital	13.0		16.9	20.9
1960 (Noren)				
Value-Added	36.5	40.7	46.7	49.9
Wages Bill	26.1	26.1	27.7	27.7
Returns to Capital	10.3	14.6	19.0	22.2

the absence of accurate information on how to allocate *other costs* among purchased inputs, wages, and capital charges, I estimate value-added exclusive of them. This measure of value-added is smaller than an alternative measure of value-added in industry which I estimate separately from official data on national income in industry less turnover taxes in industry plus amortization (Table A-4).

Both the cost account and national income estimates of value-added presented here are larger than other Western estimates of value-added in industry, notably those of Nutter (1962, p. 238) and of Noren (p. 304). The discrepancy arises mainly because the accounting capital charges in Soviet industry exceed any of the Western estimates of hypothetical capital charges. One estimate of value-added by Stanley H. Cohn (1966a p. 20) coincides with the cost account estimate presented here, but it is based on a substantially larger wages bill and smaller share of capital than is the cost account estimate.

As in the case of gross output, value-added at constant prices is estimated by deflating the current price series by the official price index. The reader hardly need be reminded of the possible statistical bias inherent in deflating value-added data by an output price index or of the probable tendency of the official price index to understate price increases. According to these estimates, constant-price value-added in industry was growing at an average annual rate of 9.1 percent from 1955–67, or slightly faster than gross output. So increasing proportions of raw materials purchases no longer contribute to the sources of upward bias in the gross output index.

III. *Factor Shares*

The share of capital and rents in industrial value-added is larger than has been estimated previously and is rising steadily during the period 1955–67. This finding is a surprising one and is at variance with the pattern of other advanced economies. In 1955, the cost account estimates allocate 68 percent of value-added to labor and 32 percent to capital and rents. By 1967, the share of labor had fallen to 47 percent; the share of capital accounted for 53 percent of value-added. These results are in contrast to Martin Weitzman's findings, which appeared too late to be analyzed here.

Viewed in the framework of the aggregate production function, the rising share of capital might be imputed to changes in the relative prices and quantities of labor and capital. But this ignores serious aggregation problems. The Soviet economy has a rapidly changing industrial structure and demonstrates a rapid rise in the relative size of industries with a high share of capital costs. Changing

aggregation is likely to account for much of the rise in the share of capital that is nominally attributed to underlying change in factor prices.

Using data for 1960 and 1964 covering 9 aggregate industrial sectors, I find that one-fourth of the rise in the overall capital-labor ratio in these sectors is due to change in the relative weights of individual industries in total value-added and three-fourths is due to the rise in capital's share in each of the individual industries.[3] Presumably, the change in factor shares within each of the 9 industrial sectors can be similarly partitioned.

Further, even within narrowly defined individual industries in the Soviet Union, there is no way to determine to what extent changes in factor shares reflect changes in underlying economic relationships or, alternatively, to what extent measured factor prices include changing amounts of implicit tax or subsidy due to changing political decisions of government planners.

In view of the aggregation problems, any estimate of the elasticity of factor substitution in an aggregate production function based on the aggregate data on changes in the relative quantities and incomes of capital and labor would be a meaningless number.

IV. *Growth of Productivity in Industry*

The sources of growth of industrial output can be estimated within the framework of an aggregate production function. Assume that output, Y, can be related to capital, K, and labor, L, at time t by an expression:

$$Y(t) = A(t)F[K(t), L(t)] \tag{1}$$

The shift parameter, $A(t)$, or total productivity, is estimated as a residual:

$$A(t) = \frac{Y(t)}{F[K(t), L(t)]} \tag{2}$$

Assume that the production function is linear homogeneous and that Euler's theorem holds. Because there is an identification problem in distinguishing between $A(t)$ and $F(K, L)$, the residual can only be estimated after the form of F is assumed.

The simplest form is the Cobb-Douglas production function with unitary elasticity of substitution, $\sigma = F_K F_L / F F_{KL}$ where the subscripts indicate partial derivatives. But the Cobb-Douglas form with fixed factor shares provides a poor approximation to the data for Soviet industry in this period when the share of capital shows a considerable increase. This can be seen by differentiating the logarithm of equation (2) with respect to t:

$$G_A = G_Y - aG_K - bG_L \tag{3}$$

where G refers to the growth rate of the subscripted variables and a and b are the imputed shares of capital and labor in income which will sum to unity in the case of constant returns to scale. With a unitary elasticity of substitution, a and b are constant. When σ is less than unity, a more rapid growth of capital than labor implies that the share of labor would increase over time. This describes the *U.S.* data.[4] When σ is greater than unity, a more rapid growth of capital than labor implies that the share of labor would fall and the share of capital would rise, as observed here in the case of postwar Soviet industry.

[3] Estimated as the geometric average of the two index numbers

$$\frac{\Sigma p_{io}s_{it}}{\Sigma p_{io}s_{io}} \quad \text{and} \quad \frac{\Sigma p_{it}s_{it}}{\Sigma p_{it}s_{io}}$$

where p_{io} measures the proportion of the ith industry in base year value-added and s_{io} measures the share of capital in the ith industry in the base year. Data are taken from my forthcoming article.

[4] These studies are surveyed by Marc Nerlove.

For the *CES* production function, Richard Nelson has demonstrated that the rate of growth of total productivity is approximately given in the two-factor case by:

$$G_A = G_Y - aG_K - bG_L - (1/2)ab\left(\frac{\sigma - 1}{\sigma}\right)G_K - G_L \tag{4}$$

When $\sigma = 1$ or when the growth rates of capital and labor are the same, (4) will be identical to (3).

In a recent issue of this *Review*, Earl Brubaker showed that his provisional calculations for the Soviet economy based on synthetic factor shares are consistent with low values of σ. He provides estimates of the growth of factor productivity in three sectors of the Soviet economy, including industry, assuming among other values an elasticity of factor substitution of 0.2. In his production function study, Weitzman found elasticities of substitution ranging in value from .274 to .403. In contrast, the data presented in this study in which a rise in the capital-labor ratio is associated with an increasing share of capital would seem to imply an elasticity of substitution of greater than unity. However, I have argued above that changing factor shares in Soviet industry reflect the changing shares of individual products in total industrial output. So the aggregate data presented here tell us little, if anything, about the elasticity of substitution. This conclusion is still more obvious in the case of estimates of σ based on synthetic factor shares, since the synthetic shares are constructed, to begin with, by applying an assumed hypothetical rate of return to capital to data on the stock of capital.

Fitting (3) to the measured growth of output and inputs provides estimates of the sources of industrial growth. The relevant data are the series on constant-price value-added, on constant-price net fixed and working capital, and on employment in industry all shown in Table 3. I use two alternative labor input series: one based on employment of production personnel in industry, the other based on employment adjusted for change in hours worked and for change in educational stock per employee, based on estimates of Stanley Cohn (1966b, p. 131).

These data yield the following average rates of growth for the period 1955–67: value-added, 9.1 percent; total capital, 9.7 percent; employment, 3.7 percent; quality adjusted labor input, 6.5 percent. The resulting measures of total productivity show a considerable range depending on the period from which the weights are taken and the measure of labor used. Using the factor shares of 1960 in Table 4, the residual attributed to total productivity grows at an average annual rate of 2.9 percent with total capital and employment inputs. The contribution of labor is increased and the growth of total productivity reduced to 1.3 percent annually when the labor input is adjusted for change in hours and estimated growth of educational stock per employee.

Although the value-added series show slightly higher rates of growth than earlier Western series based on the value of gross output, they still demonstrate the retardation that has been documented elsewhere by Norman Kaplan (p. 297). They confirm Kaplan's finding that lower growth was associated with a decline in the rate of increase in factor productivity and show that this tendency has continued in the 1960's. For the three periods 1955–59, 1959–63, and 1963–67, value-added increased at average annual rates of 10.4, 9.1, and 7.7 percent, respectively. The decline in growth of output was associated with decreases in the rate of change of measured factor productivity of 4.3, 2.8, and 1.8 percent, respectively, for the index based on total

TABLE 3—INDUSTRIAL OUTPUT AND INPUT
(Billion Constant Rubles)

	1955	1956	1957	1958	1959	1960	1961
Value-added[e]	27.95	30.79	34.04	38.52	41.43	46.02	49.94
Net fixed and working capital (Jan. 1)	52.64	59.14			81.10	87.51	96.54
Net fixed capital (Jan. 1)[f]	35.12	39.48	44.10	48.40	54.62	60.04	67.95
Working capital (Jan. 1)[g]	17.52	19.67			26.48	27.47	28.59
Employment[h] (Million workers)	18.868	19.561	20.192	20.807	21.400	22.291	23.475
Average length of the work week (1955=100)[i]	100.0	100.0	100.0	97.6	95.3	92.9	92.3
Educational stock per employee (1955=100)	100.0	105.2	110.7	116.4	119.4	122.5	125.7
Quality adjusted[j] employment (1955=100)	100.0	109.1	117.8	126.8	132.2	139.6	150.4

	1962	1963	1964	1965	1966	1967
Value-added[e]	54.08	58.74	63.50	68.15	73.43	79.17
Net fixed and working capital (Jan. 1)	106.72	116.30	128.03	139.59	151.38	160.17
Net fixed capital (Jan. 1)[f]	75.40	83.19	92.18	100.42	108.38	116.17
Working capital (Jan. 1)[g]	31.32	33.10	35.85	39.17	43.00	44.00
Employment[h] (Million workers)	24.297	25.057	25.933	27.056	28.105	28.997
Average length of the work week (1955=100)[i]	91.7	91.1	90.6	90.1	89.6	89.2
Educational stock per employee (1955=100)	129.0	132.4	135.8	139.3	142.9	146.6
Quality adjusted[j] employment (1955=100)	159.2	168.0	177.9	189.9	201.8	213.1

Notes: See Table 1.

[e] Constant-price value-added is current-price value-added deflated by the Soviet official price index and adjusted to 1955 base (Table A-6).

[f] Net fixed capital is taken to equal 75 percent of gross capital on January 1, 1960. Subsequent years are estimated by adding increment to gross capital minus amortization for replacement. *Strana Sovetov za 50 let*, p. 36; *Narodnoe Khoziaistvo SSSR* 1965, p. 64; 1964, p. 68; 1961, p. 69; *New Directions in the Soviet Economy*, II-A, p. 312. Amortization for replacement equals total amortization minus amortization for repairs. *Narodnoe Khoziaistvo SSSR*—1967, p. 883, 1965, p. 780; 1962, p. 634 (Table A-7).

[g] *Strana Sovetov za 50 let*, p. 37 (refers to December 31 of previous year). *Narodnoe Khoziaistvo SSSR 1965*, p. 761; Annual *Narodnoe Khoziaistvo SSSR* 1964, p. 751; 1963, p. 640; 1962, p. 56; 1960, p. 92.

[h] Employment from *Trud v SSSR 1968* and *Narodnoe Khoziaistvo SSSR 1965*.

[i] Data on actual time worked including overtime excluding vacations. *Vestnik statistiki*, No. 9, (1968).

[j] Stanley H. Cohn (1966b).

capital and employment and 1.7, 1.5, and 0.6 percent for the index based on total capital and quality adjusted labor (Table A-10).

This period-by-period breakdown indicates that Soviet industrial growth gradually declined from the high rates of the 1950's to rates that approached the long-run trend, estimated at 6.9 percent for the

TABLE 4—GROWTH OF TOTAL PRODUCTIVITY, 1955–67
(Average Annual Percentage Change)

Factor Shares of the Year:	Value-Added, Total Capital, Employment	Value-Added, Total Capital, Quality Adjusted Labor
1955	3.5	1.6
1960	2.9	1.3
1967	2.2	0.9

TABLE 5—GROWTH OF OUTPUT, INPUT, AND PRODUCTIVITY

Noren (1955–65)	Average Annual Percentage Change		Thornton (1955–67)
Output	8.3	8.4	Output
		9.1	Value-added
Gross fixed capital	11.1	9.7	Net fixed and working capital
Adjusted employment	3.7	3.7	Employment
Man-hours	1.6	6.5	Quality-adjusted labor
Total factor productivity (1960 weights)			Total factor productivity (1960 weights)
1) Output, gross capital, and employment	2.5	2.9	1) Value-added, net capital, and employment
2) Output, gross capital, and man-hours	4.0	1.3	2) Value-added, net capital and quality-adjusted labor

period 1928–55 in Nutter (1962 p. 163) and at 7.0–7.4 percent for 1928–58 in Powell (p. 155).

The estimates presented here may be compared with estimates by James H. Noren for the Joint Economic Committee (pp. 304 and 316). The Noren estimates are based on the period 1950–65, but I present his data for 1955–65 in Table 5 for closer comparison with this study.

Although the underlying series are altogether different—output measured as gross output by Noren, as value-added here; capital based on gross fixed capital by Noren, on net fixed plus working capital here—the implied rates of change in the two studies are not too different as long as comparisons of factor productivity are based on the simple employment series for labor. The main difference arises in the case of the adjusted employment indices. Noren adjusts the measured contribution of labor for decline in hours worked, yielding an average annual rate of increase of total productivity of 4 percent. I assume a partial offset of decreased man-hours and apply estimates by Cohn of increases in educational capital stock per employee in the whole economy to reduce the annual rate of increase of the unexplained residual to 1.3 percent.

During the course of developing these estimates, I have interjected a number of cautions and qualifications that deserve emphasis. The estimates of current price value-added are subject to error, particularly in the treatment of miscellaneous costs. The constant-price series are based on deflation by the official price index, and Soviet economists who criticised an earlier version of this study at the Central Economic Mathematical Institute in Moscow in the fall of 1969 emphasized that the official price index provides an incomplete measure of the rise in the average price of new and non-standard products. Estimates of the net capital stock are built up from estimates of gross capital and amortization which involve problems both of estimation and of deflation to constant price values. I consider the estimate of quality adjusted labor to be most tentative. However, exclusion of the contribution of increased education per worker, in turn, would introduce a real source of bias.

V. *Conclusions*

In this study, I develop estimates of Soviet industrial value-added at current and constant prices for the period 1955–67. These series provide an alternative to the available Western measures of Soviet industrial growth based on gross output data. The industrial output series presented

here show growth rates similar to those estimated by Noren for the Joint Economic Committee, but they show that value-added has increased at a faster rate than gross output in this period.

Further, the data presented here indicate that the share of capital in Soviet industrial value-added is larger than has been estimated previously and is rising during the period. Capital's share in value-added rises from .32 to .53 during the period. Part of this shift can be attributed to an increase in the relative weights of individual industries with high shares of returns to capital.

The growth of industrial inputs accounts for most of the increase in industrial output in the period. When the input index is based on net fixed and working capital and employment adjusted for growth of educational capital per employee, the average annual growth of total productivity is 1.3 percent. When the labor series are adjusted for change in human capital, the growth of total productivity in the postwar period does not appear to be particularly large.

REFERENCES

E. R. Brubaker, "Growth in Soviet Transport and Communications: Note," *Amer. Econ. Rev.*, Sept. 1969, *59*, 622–24.

———, "Synthetic Factor Shares, the Elasticity of Substitution, and the Residual in Soviet Growth," *Rev. Econ. Statist.*, forthcoming.

S. H. Cohn, (1966a) *Derivation of 1959 Value Added Weights for Originating Sectors of Soviet Gross National Product*, Research Analysis Corporation, T.P.–210 1966.

———, (1966b) "Soviet Growth Retardation," *New Direction in the Soviet Economy*, U.S. Congress, Joint Economic Committee, 1966.

E. D. Domar, "An Index Number Tournament," *Quart. J. Econ.*, May 1967, *81*, 169–88.

R. V. Greenslade, "Industrial Production in the U.S.S.R.," Conference on Soviet Statistics, Duke Univ., Nov. 1969.

N. M. Kaplan, "Retardation in Soviet Growth," *Rev. Econ. Statist.*, Aug., 1968, *50*, 293–303.

——— and R. H. Moorsteen, *Indexes of Soviet Industrial Output*, Rand Research Memorandum RM-2495, May 13, 1960.

R. R. Nelson, "The CES Production Function and Economic Growth Projections," *Rev. Econ. Statist.*, Aug., 1965, *46*, 326–28.

M. Nerlove, "Recent Empirical Studies of the CES and Related Production Functions," in M. Brown, ed., *The Theory and Empirical Analysis of Production*, New York 1967, 55–121.

J. Noren, "Soviet Industry Trends in Output, Inputs, and Productivity," *New Directions in The Soviet Economy*, U.S. Congress, Joint Economic Committee, Washington 1966, 271–326.

G. W. Nutter, *Growth of Industrial Production in the Soviet Union*, Princeton 1962.

———, "Industrial Growth in the Soviet Union," *Amer. Econ. Rev.*, May 1958, *48*, 398–411.

R. P. Powell, "Industrial Production," in A. Bergson and S. Kuznets, eds., *Economic Trends in the Soviet Union*, Cambridge 1963, 150–202.

T. V. Riabushkin, *Ekonomicheskaia statistika*, Moscow 1966.

S. G. Stoliarov, *Otsenakh i tsenoobiazovanii v SSSR*, Moscow 1963.

J. Thornton, "The Estimation of Value-Added and Average Returns to Capital in Soviet Industry from Cross-Section Data," *J. Polit. Econ.*, Dec. 1965, *73*, 620–35.

———, "Differential Capital Charges and Resource Allocation in Soviet Industry," *J. Polit. Econ.*, forthcoming.

M. L. Weitzman, "Soviet Postwar Economic Growth and Capital-Labor Substitution," *Amer. Econ. Rev.*, Sept. 1970, *60*, 676–92.

U.S. Congress, Joint Economic Committee, *Soviet Economic Performance: 1966–67*, Washington 1968.

Changes in the Nonwhite/White Income Ratio—1939-67

By JAMES GWARTNEY*

Public concern about color discrimination has led to increased interest in the measurement of changes in the nonwhite/white income ratio (*NWIR*).[1] If there has been a reduction in effective employment discrimination against nonwhites in recent years, we would expect this ratio to increase, other things held constant. However, other things are not constant. If one is interested in estimating changes in the intensity of discrimination, particularly employment discrimination, it will be necessary to disaggregate income data in an effort to isolate other factors influencing the *NWIR*.

The purpose of this paper is to estimate changes in the *NWIR* between 1939 and 1967 for persons 25 years old and over, and to analyze the impact of changes in regional composition, scholastic achievement, quantity of education, and structural demand for highly educated labor on the magnitude of the white-nonwhite income differential during this period.

Sections I and II contain aggregated and disaggregated estimates of changes in the *NWIR* for the 1939–67 period according to sex. Section III is an analysis of employment discrimination and other factors that influence the *NWIR*. Section IV contains comments about changes in the intensity of discrimination in light of the data.

Briefly, the estimates indicate the *NWIR* for males increased substantially—probably more than 12 percent *within regions*—during the 1940's. A comparable figure for the 1949–67 period is a 1 to 2 percent increase in the *NWIR*. If mean rather than median income data are used, the ratio increased between 5 and 7 percent within regions in the latter period.

After increases of between 2 and 4 percent during the 1940's, the income of nonwhite females has increased very rapidly—an estimated 25 to 30 percent relative to white females even after adjustment for migration—during the 1949–67 period.

While migration has contributed to income equality according to color, changes in other variables have reduced income gains of nonwhites. I shall argue below that the structure of recent income gains and the changes in scholastic achievement differentials between whites and nonwhites have each retarded the relative income growth of nonwhite males an estimated 3 to 5 percent during the 1949–67 period, largely offsetting the positive impact of migration.

I. *Education, Migration, and Changes in the NWIR for Males 25 years old and over—1939–67*

Mean wage and salary data for urban males 25 years of age and over, obtained from the 1940 Census, have been standardized according to age, education, region and color for 1939 by Morton Zeman. Comparable data for mean earnings for non-farm occupations can be derived for

* Assistant professor of economics at Florida State University. I wish to express appreciation to Professors Kenneth M. McCaffree, Al Holtmann, William E. Laird, David Rasmussen, and Howard Tuckman for helpful comments on an earlier draft of this paper. The research assistance of P. Hanumantha Rayappa is also gratefully acknowledged. The research described in this paper was partially supported by a grant from the Manpower Administration, U.S. Department of Labor.

[1] See Alan Batchelder, Gary Becker, Rashi Fein, and David Rasmussen.

1959 from the 1960 Census. Zeman's results indicated the mean wage and salary income of nonwhites 25 to 64 years of age was 58.2 percent of whites for urban males in the North and West and 40.5 percent in the South in 1939. In 1959 the comparable mean earnings of nonwhites were 63.7 percent of whites in the North and West and 47.1 percent in the South for males in nonfarm occupations. Thus, the relative mean income of nonwhite urban males, age 25 to 64, increased by 16.2 percent in the North and West and 21.7 percent in the South relative to whites during the 20-year period.

Table 1 indicates the results of disaggregating the total income change of urban males into an "income effect" and a "distribution effect" for the 1939–59 period. The income effect is the change in income assuming the education and age composition of a population remained constant.[2] The distribution effect estimates the contribution of changes in the educational and age composition of a population between periods on changes in total money income, holding the income *within* education-age cells constant.[3] While nonwhites increased their income an estimated 23 to 25 percent within education-age cells during the period, the influence of the distribution effect was negative, partially negating the relative income gains of nonwhites with similar quantity of education as whites. Equations (2) and (2′) shown in footnote 3 estimate that the distribution effect *reduced* the income of nonwhites relative to whites between 5.8 and 7.3 percent in the North and West and 3.6 and 3.8 percent in the South during the period.[4]

The distribution effects of education and age were negative between 1940 and 1960 because the quantity of education gains of nonwhites relative to whites was small, and the income differential was greater in higher education cells during the period. The latter point is important because as the general level of education increased, nonwhites moved into education cells with a larger white/nonwhite income differential. The data of James Coleman indicate the white/nonwhite scholastic achievement differential is greater in higher educational categories. Thus the larger income differential in higher education cells is, at least partially, a reflection of the

[2] The mathematical form used to estimate the income effect is:

(1) $$\frac{\Sigma(Y^*_2 \cdot E_1)}{\Sigma(Y^*_1 \cdot E_1)} \quad \text{(Laspeyres)}$$

and,

(1′) $$\frac{\Sigma(Y^*_2 \cdot E_2)}{\Sigma(Y^*_1 \cdot E_2)} \quad \text{(Paasche)}$$

Where Y^* is the income within an education-age cell of those 25 years old and over with income. E is the percentage of a population with income within the education-age cell, and, subscripts 1 and 2 indicate time periods, the more recent time period represented by 2.

The above formula can be applied, subject to data limitations, to education-age cells according to color and region. The income effect indicates the contribution of changes in income *within* education-age cells, holding the education-age distribution of a population constant —first at the distribution of period one (Equation 1), then period 2 (Equation 1′). If the income within education-age cells increases between period 1 and period 2, the value of the income effect will be greater than 1. When comparing two populations, the population with the more (less) rapid gain between periods will have increased (decreased) its relative income.

[3] The form of the distribution effect is:

(2) $$\frac{\Sigma(Y^*_2 \cdot E_2)}{\Sigma(Y^*_2 \cdot E_1)} \quad \text{(Laspeyres)}$$

and

(2′) $$\frac{\Sigma(Y^*_1 \cdot E_2)}{\Sigma(Y^*_1 \cdot E_1)} \quad \text{(Paasche)}$$

The income effect multiplied by the distribution effect equals the "total" effect. In the case of mean income data, the total effect will equal the income in period 2 divided by income in period 1.

[4] A rerun of the data adjusting for first, only age, and second, only education, indicates changes in the educational distribution are responsible for the negative distribution effect. Changes in the age distribution of the two populations had virtually no impact on the nonwhite/white income ratio during the period.

greater scholastic achievement differential for these educational classes. Additional discussion of the scholastic achievement factor is contained in Section III of this paper.

Between 1940 and 1960, the percent of nonwhite urban males residing in the South decreased from 58.2 to 50.8. Table 1 indicates their mean income is estimated to have increased between 3.1 and 3.7 percent relative to whites because of this change in regional composition.[5]

Recently, considerable discussion has centered on the 1949–59 period. Median and mean income data standardized according to education, region, color, and sex can be derived from the 1950 and 1960 Census. Since the data are not standardized for urban-rural or age composition, direct comparison with 1939–59 data is not possible.[6] However, the data are useful in isolating income changes according to color during the 1950's.

Table 1 presents estimates of factors influencing median and mean income changes by region for males during the 1950's. For mean incomes, the *NWIR* of males 25 and over increased from 49.3 in 1949 to 51.5 in 1959—an increase of 4.6 percent during the period.[7] However, within regions the *NWIR* for males decreased slightly both in the North and the South. When median income data are used, the reduction in the *NWIR* during the 1950's is slightly greater. Table 1 also indicates the results of disaggregation for the 1949–59 period. Nonwhite males failed to increase their income relative to whites with a similar quantity of education and regional location.

The *Current Population Reports* (*CPR*) contain mean and median income data for the United States according to education and race for those 25 years old and over for 1967.[8] These data can be compared with the census data, estimating recent income changes according to education, color, and sex.

For mean incomes of males, the *NWIR* increased from 51.5 percent in 1959 to 57.4 percent in 1967—an increase of 11.6 percent during the period (see Table 1). Using median data, the relative income of nonwhite males was 53.9 percent of whites in 1959 and 58.2 percent in 1967—an increase of 8 percent. These increases were largely the result of increases in income within education cells. The education effect was small in the case of median data and negative when mean income data were

[5] Migration from the South to the North increases the relative income of nonwhite in the United States because incomes in the North for those with similar education are higher than in the South, and the incomes of similarly educated nonwhites relative to whites are greater in the North. The North-South income differential for nonwhites is used to approximate the impact of migration on the income of nonwhites. The form of the migration effect is:

$$\frac{\Sigma(Y^N \cdot E^{nw})}{\Sigma(Y^S \cdot E^{nw})} \cdot M$$

where, Y is the mean income of nonwhites 25 and over within an education-age cell in the base year.
E^{nw} is the percentage of nonwhites within an education-age cell for the United States.
N and S are subscripts representing North and South, and
M is the change in the percentage of the nonwhite urban population residing in the North during the period.

Estimates can be made using income data for both earlier and later periods.

[6] Since the distributional changes are for education only, they will be referred to as the education effect. Because the age distributions of the two populations change slowly, age has little impact on changes in relative income.

[7] In all cases the midpoints of class intervals were assumed to be the mean of the class. For the open-end classes, $20,000 was assumed to be the mean of the $10,000 open-end class for 1949 data and $30,000 was used for the mean of the $15,000 open-end class in 1969. These means were considered appropriate in light of their use in the past by W. L. Hansen and Herman Miller and after investigation of other 1960 Census data.

[8] The *CPR* data are for Negroes, rather than nonwhites. However, since 95 percent of all nonwhites are Negroes, and since the income of other nonwhites, exclusive of Oriental Americans, are similar to those of Negroes, one would expect the white/Negro income differential to be almost identical to the white/nonwhite income differential.

TABLE 1—INCOME, DISTRIBUTION, AND MIGRATION EFFECTS ON THE NONWHITE/WHITE INCOME RATIO OF MALES 25 YEARS OF AGE AND OVER FOR 1949–59, 1959–67, AND FOR URBAN MALES 1939–59

	Percent Change in the Nonwhite/White Income Ratio						
	1939–59 (Urban)			1949–59			1959–67
Mean Income	North	South	United States	North	South	United States	United States
Actual (NW/W)	16.2	21.7	22.8	−0.2	−2.7	4.6	11.6
Total Effect (NW/W)	16.2	21.7	22.8	−0.2	−2.7	4.6	11.6
Income Effect (NW/W)							
Laspeyres	23.3	26.9	23.1	0.9	−1.8	2.1	13.2
Paasche	25.3	27.3	25.2	1.3	−0.8	2.6	13.3
Distribution Effect (NW/W)							
Laspeyres	−5.8	−3.6	−0.2	−1.1	−1.0	2.6	−1.5
Paasche	−7.3	−3.8	−1.9	−1.6	−1.9	2.1	−1.4
Migration Effect (NW/W)							
Laspeyres	—	—	3.1	—	—	5.1	3.3
Paasche	—	—	3.7	—	—	5.7	3.5
Median Income							
Actual (NW/W)	—	—	—	−1.9	−6.7	3.1	8.0
Total Effect (NW/W)	—	—	—	−1.3	−5.3	3.8	7.2
Income Effect							
Laspeyres	—	—	—	−0.6	−6.2	−1.2	7.0
Paasche	—	—	—	−0.1	−5.3	−0.7	6.3
Distribution Effect (NW/W)							
Laspeyres	—	—	—	−1.1	0.9	5.0	0.1
Paasche	—	—	—	−1.6	0.1	4.5	0.8
Migration Effect (NW/W)							
Laspeyres	—	—	—	—	—	5.6	3.6
Paasche	—	—	—	—	—	6.7	3.9

Source: The 1939 mean income data for urban males were derived from Zemen. The comparable 1959 data are for non-farm occupations and were derived from *U.S. Census of Population: 1960*, Subject Report, "Occupation by Earnings and Education." The mean and median income data used for all males were derived from *U.S. Census of Population: 1950*, Special Report, "Education," *U.S. Census of Population: 1960*, Subject Report, "Education Attainment," and *Current Population Reports, Series P-60*. The distribution effect for 1949–59 and 1959–67 is really an education effect only, as data are unavailable for age-education cells.

used. Like the 1950's, changes in the regional composition had a positive impact, increasing the income of nonwhite males between 3 and 4 percent relative to whites during the 8-year period.

If we adjust initially for the regional composition factor, the results would suggest that for median incomes of males 25 and over, the *NWIR* increased substantially during the 1940's, and remained relatively stable during 1949–67.[9] While

[9] Since the income of nonwhite males is more favorably affected by cyclical expansion than that of white males, both the estimated relative income gains of males for the 1940's and the 1959–67 period are biased upward, while those for the 1950's are biased downward. Rasmussen estimated the importance of cyclical variables on the nonwhite/white income ratio for the post war period. Using 1948–64 time-series data, Rasmussen regressed the nonwhite/white median income ratio (Y_t) on the rate of growth of GNP ($\%\Delta GNP$), the unemployment rate lagged one period (U_{t-1}) and a time trend (T) and found:

$$Y_t = .57 + \underset{(.0025)}{.00984}(\%\Delta GNP) - \underset{(.0077)}{.0270}(U_{t-1}) + \underset{(.0015)}{.00323}(T) \qquad R^2 = .57 \quad df = 13$$

Both cyclical variables have the expected sign and are significant at the .02 level. Substituting into Rasmussen's equation, data indicate the change in the actual nonwhite/white income ratio for the 1950's is understated by 1.7 percent. The actual change in the nonwhite/white ratio is biased upward an estimated 4.4 percent for the 1959–67 period and by 2.5 percent for

the male *NWIR* increased by 11.2 percent during the recent 18-year period,[10] the regional composition factor is estimated to have increased between 9.2 and 10.9 percent during the period.[11] Therefore, almost all of the relative median income gains of nonwhites during the period are estimated to have resulted from regional migration.

If mean rather than median income data are used, the relative income gains of nonwhites will be greater, and the net change in male *NWIR* after adjustment for migration was found to be 5 to 7 percent during the 1949–67 period.

II. *Education, Migration, and Changes in the NWIR for Females 25 years old and over—1939–67*

Mean income data, similar to the 1939–59 data for urban males presented in Table 1, are unavailable for females. However, changes in the relative mean and median income ratio for 1949–59 and 1959–67 can be derived. These estimates are presented in Table 2. The results confirm a rapid increase in the *NWIR* of females between 1949 and 1967.

Using mean income data, Table 2 suggests the relative income of nonwhite females increased 13.0 percent in the North and 16.2 percent in the South during the 1950's. For the United States, using mean incomes, the *NWIR* is estimated to have increased 20.9 percent during the 1950's, and between 14 and 15 percent after adjustment for regional migration. Using median income data, the female *NWIR* increased by 11.0 percent for the United States and 14.6 and 13.5 percent in the North and South, respectively.[12] These increases were due to both income and education effects during the 1949–59 period.

During the 1960's, the *NWIR* of females continued to increase rapidly. The median income of nonwhite females increased an estimated 68.5 percent compared to 33.2 percent for white females—an increase of 26.5 percent in the actual *NWIR*.[13] When the entire 1949–67 period is considered, the gains of nonwhite females relative to whites are quite impressive. Using mean or median income data, the *NWIR* increased between 25 and 30 percent even after adjustment for changes in the regional composition factor during the 18-year period.[14]

the entire 1949–67 period. Annual unemployment rates for 1949 and 1967 were identical, 3.8 percent, but growth of *GNP* was greater in 1967 and 1949.

[10] The median income increase was 163.5 and 136.9 percent for nonwhite and white males, respectively.

[11] The migration effect was again estimated by the method of fn. 5. The percent of the nonwhite population residing in the North increased by 15.8 percent for males and 14.1 percent for females during the 18-year period.

[12] The total income effect for the United States was much greater than the actual nonwhite/white increase in median income during the 1950's, even though they were similar within regions. When using median income data, discrepancies result from differences in the shapes of the income distributions of the two populations between periods and regions. It should be noted both mean and median income data yield similar estimates of the actual and total effect within regions during the 1950's and for the United States during the entire 1949–67 period.

Due to data limitation, the distribution changes of this section are for education only, and they will be referred to as an education effect.

[13] The percent increase in median income according to color and sex is given below:

	1939–49	1949–59	1959–67	1949–67
Nonwhite males	197.2	76.0	49.8	163.5
White males	146.0	70.7	38.7	136.9
Nonwhite females	165.8	50.2	68.5	153.2
White females	138.9	36.6	33.2	81.5

All data are median income for those 25 and over, except for the 1939–49 period. Data for that period are median wage and salary income for those 14 and over from the *CPR* survey. The 1967 data are also from *CPR*. Data for 1949 and 1959 are from the *Census*.

[14] In contrast to its impact in the case of males, cyclical expansion is negatively correlated with the nonwhite/income ratio of females. Apparently during an expansionary period, the nonwhite family unit is less dependent on the income of the female and she is more likely to be a part-time worker. But during a recessionary

TABLE 2—INCOME, EDUCATION, AND MIGRATION EFFECTS ON THE NONWHITE/WHITE INCOME RATIO FOR FEMALES 25 YEARS OF AGE AND OVER FOR 1949–59 AND 1959–67

	Percentage Change in Nonwhite/White Income Ratio 1949–59			1959–67
Mean Income	North	South	United States	United States
Actual (NW/W)	13.0	16.2	20.9	14.1
Total Effect (NW/W)	13.0	16.2	20.9	15.1
Income Effect (NW/W)				
Laspeyres	7.8	9.6	10.8	11.7
Paasche	9.9	10.0	12.0	12.9
Education Effect (NW/W)				
Laspeyres	4.8	6.0	9.0	3.0
Paasche	2.8	5.6	7.9	2.0
Migration Effect (NW/W)				
Laspeyres	—	—	5.7	3.4
Paasche	—	—	6.6	4.0
Median Income				
Actual (NW/W)	14.6	13.5	11.0	26.5
Total Effect (NW/W)	14.2	12.3	21.5	18.9
Income Effect (NW/W)				
Laspeyres	7.4	6.4	9.5	16.0
Paasche	10.2	5.6	10.4	17.2
Education Effect (NW/W)				
Laspeyres	6.4	5.5	11.0	2.5
Paasche	3.6	6.3	10.0	1.5
Migration Effect (NW/W)				
Laspeyres	—	—	5.9	4.4
Paasche	—	—	6.8	4.6

Source: See Table 1.

The failure of the money income of white females to increase as rapidly as the other three color/sex groupings is largely responsible for the substantial relative income gains of nonwhite females. Only during the 1960's has the median income of nonwhite females increased relative to males. There was little difference in the rate of increase in the median income of nonwhite males, white males, and nonwhite females during the entire 1949–67 period, after adjustment for the regional migration factor.

period, the nonwhite female is more likely than her white counterpart to be forced into full-time employment. Therefore, the estimated relative income gains of nonwhite females during the 1940's and 1960's are downward biased and during the 1950's the estimates are upward biased. The direction of these biases at least partially explain why Batchelder found the relative income of nonwhite females, but not males, increased during the 1950's. Considering the cyclical factor, the income gains of nonwhite females during the 1960's are even more impressive.

III. *Employment Discrimination and Changes in the NWIR*

The *NWIR* at any point in time is a function not only of current employment discrimination but also past employment discrimination, and discrimination, both current and past, in the acquisition of productivity factors. In 1959, between two-fifths and two-thirds of the white/nonwhite income differential was estimated to result from differences in productive capacity.[15] Even if the relative productive

[15] See my article in this *Review* for the derivation of these estimates.

capacity of nonwhites entering the labor force should improve due to more favorable opportunities to acquire human capital and other productivity factors, the *NWIR* would improve only slowly. Each year new entrants account for approximately 3 percent of the total labor force. Even after a decade, those who acquired most of their productive capacity during the immediate 10-year period would compose only one-fourth to one-third of the total labor force. The impact of past discrimination on present productive capacity, and, consequently, present earnings differences, indicates why attainment of equality in income between races is a slow process.

If however, the intensity of employment discrimination is declining, we would expect *ceteris paribus*, the nonwhite/white income ratio to increase. But in the real world, variables other than employment discrimination are not constant. Changes in the intensity of employment discrimination cannot be isolated without first isolating changes that are essentially unrelated to discriminatory behavior in employment. We proceed to consider a number of these factors.

The results of Sections I and II suggest that the migration of nonwhites from the South to the North has been an important source of growth in the relative income of both nonwhite males and females. It is estimated that the migration factor contributed between 7 and 9 percent to the male *NWIR* during the 1940's; 5 to 6 percent during the 1950's, and that this positive impact has continued during the 1960's. Similar estimates were obtained for females.

While the migration factor has positively influenced the *NWIR*, this section presents evidence suggesting that the structure of recent income gains has been unfavorable to nonwhites and that moreover the structure of the white/nonwhite scholastic achievement differential results in a widening achievement gap between the two populations as the general level of education increases. These two factors, not directly related to employment discrimination, have reduced the relative income gains of nonwhites in the most recent years. Their combined negative effect has largely offset the positive impact of migration.

During recent years the more rapid income gains, in the case of both whites and nonwhites, have been made by those in the higher educational classes.[16] However, since whites are overrepresented relative to nonwhites among the well educated, the structure of income growth has reduced the relative income gains of nonwhites.

The impact of this factor can be estimated by comparing the income effect of nonwhites (equations (1) and (1′), footnote 2) with a hypothetical income effect, assuming nonwhites had the white educational distribution.[17] The results of that comparison indicate the income *gains* of nonwhite males would have been between 3.1 and 3.5 percent greater during the 1950's if their educational distribution had been identical with that of whites (Table 3). During both the 1950's and 1960's, in the North as well as the South, the hypothetical income effect was greater than the actual for both males and females, indicating the greatest income gains were made by

[16] See Miller for evidence that the rate of mean income growth during the 1950's was positively related to quantity of education. While the results given in this section are for mean income, the same transformations were performed and similar results obtained using median income data.

[17] The mathematical form of the Paasche hypothetical nonwhite income effect is:

$$\Sigma(Y_2^n \cdot E_2^w)/\Sigma(Y_1^n \cdot E_2^w)$$

The hypothetical income effect (using the white educational distribution) will be greater than the nonwhite income effect (using the nonwhite educational distribution) if larger income gains are made by those with more education. Comparison can be made between both the Laspeyres and Paasche indexes.

those with the most education. Only in urban areas during the 1939–59 period was the actual nonwhite income effect greater than the hypothetical.

Dale Hiestand has indicated that during the 1940's there was a tendency for wage rates in occupations at the bottom of the income scale to increase more rapidly than those at the top. There was also a narrowing of income differentials between the well educated and those with less education. Since nonwhites are overrepresented among those with less education, the structure of income gains during the 1940's, unlike the 1950's and 60's, was favorable to them, partially accounting for the larger income gains during the period.

In addition to the differential impact of income gains among education classes, the structure of scholastic achievement differences between the two populations has retarded growth of the nonwhite/white income ratio. Even though the educational gap between whites and nonwhites has narrowed slightly in terms of quantity of education, the education effect for males has either been negative or only slightly positive during each period considered by this paper. The income gap between white and nonwhite males is greater, in both absolute and relative terms, within higher educational categories. This is, at least partially, the result of a larger scholastic achievement gap between whites and nonwhites in higher education cells.[18] Therefore, *ceteris paribus*, the achievement gap increases with the general educational level of the two populations, even if the differential is constant at each grade level over time. Of course this factor could be partially or entirely offset by increases in the scholastic achievement of nonwhites relative to whites within education cells.[19] Increased expenditures on nonwhite education relative to white, and Supreme Court decisions on desegregation suggest a possible narrowing of achievement differentials in recent years inasmuch as they are related to differences in the quality of schools, but there is little direct evidence for this hypothesis. In any case, the larger achievement gap in higher education cells would tend to reduce the impact of any reduction in the scholastic achievement within cells, as the general level of education of both populations increases over time.

In addition to the greater scholastic achievement gap in higher educational categories, there is some evidence to suggest that employment opportunities for non-

TABLE 3—ESTIMATED IMPACT OF THE STRUCTURE OF MEAN INCOME GAINS ON THE NONWHITE/WHITE MEAN INCOME RATIO FOR VARIOUS TIME PERIODS

Time and Region	Hypothetical Nonwhite Income Effect (assuming the white education distribution) divided by the Actual Nonwhite Income Effect			
	Males		Females	
	Laspeyres	Paasche	Laspeyres	Paasche
1939–59				
United States	.973	.961	—	—
North	.988	.987	—	—
South	.976	.964	—	—
1949–59				
United States	1.035	1.031	1.035	1.022
North	1.031	1.026	1.048	1.028
South	1.046	1.043	1.022	1.016
1959–67				
United States	1.014	1.019	1.017	1.013

Source: See Table 1.

[18] A study of the U.S. Office of Education, the *Coleman Report*, estimated scholastic achievement differences between the two populations in terms of quantity of education (years of schooling) for three different grade levels in both metropolitan and nonmetropolitan areas. In each case the results indicated a greater scholastic achievement differential in the higher education cells.

[19] However, the point is that the scholastic achievement gap increases, not because of changes in the scholastic achievement differential within an education (quantity) cell, but because the differential is greater in higher educational cells, which increases in importance as the general level of education increases. Lowell Gallaway, among others, has recognized the potential of this factor to adversely affect the nonwhite/white income ratio over time.

TABLE 4—ESTIMATED IMPACT OF THE STRUCTURE OF SCHOLASTIC ACHIEVEMENT DIFFERENCES BETWEEN WHITES AND NONWHITES ON THE NONWHITE/WHITE MEAN INCOME RATIO FOR VARIOUS TIME PERIODS

Time Period and Region	Hypothetical Nonwhite Education Effect (assuming the white rate of change in income per unit of education) Divided by the Actual Nonwhite Education Effect			
	Males		Females	
	Laspeyres	Paasche	Laspeyres	Paasche
1939–59				
United States	.984	1.022	—	—
North	1.057	1.079	—	—
South	1.028	1.050	—	—
1949–59				
United States	1.028	1.019	.998	.989
North	1.036	1.032	1.014	1.019
South	1.041	1.047	1.032	1.008
1959–67				
United States	1.014	1.002	.985	.996

Source: See Table 1.

whites, particularly nonwhite males, are less favorable for occupations requiring higher skill and educational levels. The increasing scholastic achievement gap, possibly reinforced by the occupational pattern of discrimination, adversely affects the ability of nonwhites, particularly males, to increase their income relative to whites because these factors result in larger white/nonwhite income differentials as quantity of education increases.

The impact of the structure of scholastic achievement differences can be estimated quantitatively by comparing the education effect (equation 2 and 2′, footnote 3) of nonwhites with an hypothetical education effect, assuming nonwhites had the same rate of increase in income as quantity of education increases, as in the case of whites.[20] Table 4 indicates the income of both nonwhite males and females would have increased more rapidly within regions during the 1950's, given their educational gains, if their rate of income increase per unit of education had been similar to that of whites. Within regions, this factor is estimated to have reduced nonwhite mean income gains from 3 to 5 percent for males and 1 to 3 percent in the case of females during the 1949–59 period. During the 1939–59 period, the mean income gains of nonwhite males were reduced an estimated 2.8 to 5.0 percent in the South and 5.7 to 7.9 percent in the North because their income did not increase as rapidly as whites for similar gains in quantity of education. The estimates would indicate the factor is somewhat less important during the most recent 1959–67 period but this conclusion must be tentative since regionally adjusted data for this period are unavailable. As indicated above, the slower rate of increase in income for nonwhites, as quantity of education increases, may not be entirely attributable to the larger scholastic differentials in higher education cells. However, the scholastic achievement factor must be a major component contributing to this reduced rate.[21]

John Kain and Joseph Mooney[22] have suggested the postwar suburbanization of employment opportunities has adversely influenced the income of nonwhites since they reside predominately in the central

[20] The mathematical form of the hypothetical. Laspeyres education effect is: $\Sigma\,(Y_2^w \cdot E_2^n)/\Sigma(Y_2^w \cdot E^n_1)$. Since the rate of increase in income as education increases is greater for whites than nonwhites, the hypothetical income effect will be greater than the actual nonwhite income effect. A Paasche index, using the white income within education cells for period 1, can be derived, also.

[21] If discrimination is less intense against those in higher educational categories, the estimates will understate rather than overstate the impact of the structure of scholastic achievement differences on income change over time. The evidence would indicate this is more likely to be the case for nonwhite females than males.

[22] Mooney also argues that the growing employment sectors in the central cities such as service, finance, insurance, and real estate are largely employers of females. Changes in the composition of jobs within central cities, while unfavorable to nonwhite males aspiring a high wage blue collar job in manufacturing, are favorable to females—both white and black. This is consistent with the more rapid relative income gains of nonwhite females in recent years.

areas of cities. Considering the negative impact of this factor, plus those previously discussed in this section, in recent years we might have expected the relative income of nonwhites, regionally adjusted, to have fallen in absence of a decline in the intensity of employment discrimination. However, this was not the case. During the 18-year period from 1949 to 1967, for mean incomes the *NWIR* for males, regionally adjusted, increased an estimated 5 to 7 percent. While median income gains were less, between 1 and 2 percent relative to whites, they were still positive. In urban areas, from 1939 to 1959, the income of similarly educated (quantity) nonwhite males increased from estimated 23 to 27 percent relative to whites within regions. The magnitude of the relative income gains of nonwhite urban males during the 1939–59 period and the recent income gains that took place in spite of other unfavorable factors, suggest a decline, albeit a small one, in the intensity of employment discrimination against nonwhite males.

The evidence indicating a decline in employment discrimination appears to be even stronger for nonwhite females. Since 1949, the *NWIR* of females has increased an estimated 25 to 30 percent, even after adjustment for migration. This rapid increase in income was accomplished despite an increasing labor force participation rate and the negative influence of the structure of income gains and scholastic achievement differentials.

Between 1939 and 1967 for median female incomes the *NWIR* more than doubled.[23] Nearly one-half of the median income differential present in 1949 had been eliminated by 1967. In the North, according to the *CPR Survey*, by 1967 the income of nonwhite females in the labor force was estimated to be greater than for white females. During the 1960's despite the negative influence of economic expansion (see footnote 14) and their increased labor force participation, the income of nonwhite females has increased more rapidly than any other color/sex grouping.

The income of nonwhite females has increased due to both the income and education effects (Table 2). Unlike males, the income differentials between nonwhite and white females is not greatest in the highest education cells.[24] This is at least partially because of favorable employment opportunities in professional occupations such as teaching and nursing.[25] If the highly educated nonwhite females have relatively better employment opportunities than is the case for those with less education, this factor could offset the greater sholastic achievement gap in higher education cells. The fact that the relative education effect was usually positive for females but negative for males would indicate the existence of more favorable employment opportunities for highly educated nonwhite females. While reduced income differentials resulting from a decline in employment discrimination are difficult to isolate, the income changes of nonwhite females since 1949 are certainly consistent with such a reduction in discrimination.[26]

[23] The nonwhite/white income ratio of females for various years is indicated below:

	1939	1949	1959	1967
Median Income Ratio (NW/W)	.364	.536	.589	.745

All data are for those with income, 25 years of age and over, except the 1939 figure. The 1939 data are for those with income, 14 years of age and over.

[24] This is particularly true for nonwhite females with a college education—i.e., the white/nonwhite income differential is small in this education cell for females.

[25] Seventy-seven percent of all nonwhite females in professional occupations are either teachers or nurses—two low discrimination occupations. See Marshall Colberg for additional evidence that discrimination against nonwhites is low in teaching, particularly in the South.

[26] Other hypotheses are consistent with the increase in nonwhite/white income ratio for females. For example, the role of the white female in the labor market is considerably different than the nonwhite. The white female is more likely to be a secondary worker. As family incomes have increased in recent years white females may

IV. *The Future Direction of Relative Income of Males*

Since the male is most often the primary source of income in our society, future income changes of nonwhite males are particularly important. The data of this paper suggest that future relative income gains of nonwhite males are likely to be slow even if there is some reduction in employment discrimination.

First, the impact of the past cannot be ignored. A substantial portion of the white/nonwhite income differential is the result of present productivity differences between the two populations. These productivity differences change slowly.

Second, the migration factor, a strong positive influence in the past, cannot be expected to exert as large an influence on the relative income gains of nonwhites in the future. The migration during the 1940's and 50's was largely because of movement from South to North and rural to urban. Since two-thirds of the southern nonwhite population is now urban, compared with one-third in 1940, future migration to the North is more likely to be from the urban South. As can be seen by the data of Table 1, migration is less potent when it is from the urban South to the urban North. In addition, there is some evidence of a reduction in the rate of migration from the South to the North. The percent of the nonwhite population residing in the South decreased by 1.4 percent between 1964 and 1968, compared with 5.6 percent between 1960 and 1964, and 8 percent during the decade of the 1950's.

The structure of income gains and scholastic achievement differentials has not only retarded recent income gains of nonwhites—particularly nonwhite males, but they also paint a pessimistic picture of the future for those interested in income equality according to race. The impact of the structure of scholastic achievement differences, no doubt partially the result of cultural and environmental factors, is not likely to change in the near future. This factor will continue to increase the difficulty of obtaining achievement equality even as quantity of education differences are reduced.

The future influence of the structure of income gains cannot be predicted with certainty. It is possible that increases in the supply of highly educated personnel may dominate and the structure of future income gains, as during the 1940's, may exert a favorable influence on the *NWIR*. However, it is probably more reasonable to assume that the demand for low skilled labor generated by the wartime economy of the 1940's was the atypical case. And in the 1970's, as during the 50's and 60's it seems more probable to expect that technological change and other demand factors will favorably influence the demand for highly educated employees, and as the result, continue to adversely influence the nonwhite/white income ratio.

The one bright indicator is the quantity of education differential. Between 1940 and 1960, quantity of education differences between white and nonwhite males in the labor force over 25 years of age decreased only slightly, from 3.3 to 3.1 years. However, by 1980 the differential will be reduced to less than 2.0 years. This sharp reduction will contribute to income equality according to color. However, given the anticipated negative influence of other factors, we can expect the relative income gains of nonwhite males during the 1970's will be modest, perhaps even slower than during the 1960's.

very well have placed more emphasis on nonpecuniary job working conditions and leisure since such services are likely to have an income elasticity of more than unity. Thus, the reduction in relative income of white females may be only in money terms. In this instance, money income is a poor measure of the "total" welfare of white females.

REFERENCES

A. Batchelder, "Decline in the Relative Income of Negro Men," *Quart. J. Econ.*, Nov. 1964, *78*, 525–48.

G. Becker, *The Economics of Discrimination*, Chicago 1957.

M. R. Colberg, *Human Capital in Southern Development—1939–63*, Chapel Hill 1965.

J. S. Coleman, et al., *Equality of Education Opportunity*, Washington 1966.

R. Fein, "Relative Income of Negro Men, Some Recent Data." *Quart. J. Econ.*, May 1966, *80*, 336.

L. E. Gallaway, "The Negro and Poverty," *J. Bus., Univ. Chicago*, Jan. 1967, *40*, 27–35.

J. D. Gwartney, "Employment Discrimination, Productivity Factors and Income Differentials between Whites and Nonwhites," *Amer. Econ. Rev.*, June 1970, *60*, 396–408.

W. L. Hansen, "Total and Private Rates of Return to Investment in Schooling," *J. Polit. Econ.*, Apr. 1963, 128–40.

D. L. Hiestand, *Economic Growth and Employment Opportunities for Minorities*, New York and London, 1964.

J. F. Kain, "Housing Segregation Negro Employment, and Metropolitan Decentralization," *Quart. J. Econ.*, May 1968, *82*, 175–97.

H. P. Miller, "Lifetime Income and Economic Growth," *Amer. Econ. Rev.*, Sept. 1965, *55*, 834–44.

J. D. Mooney, "Housing Segregation, Negro Employment, and Metropolitan Decentralization: An Alternative Perspective," *Quart. J. Econ.*, May 1969, *83*, 299–311.

D. W. Rasmussen, "A Note on the Relative Income of Nonwhite Men, 1948–64," *Quart. J. Econ.*, Feb. 1970, *84*, 168–72.

M. Zeman, "A Quantitative Analysis of White-Nonwhite Income Differentials in the United States in 1939," unpublished doctoral dissertation, Univ. Chicago, 1955.

U.S. Census of Population: 1950, *Special Report*, "Education."

———: 1960, *U.S. Summary, Final Report*, "Characteristics of the Population."

———, *Subject Report*, "Educational Attainment," PC(2)-5B.

———, *Subject Report*, "Occupation by Earnings and Education," PC(2)-7B.

U.S. Department of Commerce, *Current Population Reports*, P-60 Series.

Durability of Consumption Goods

By PETER L. SWAN*

In a recent article in this *Review*, David Levhari and T. N. Srinivasan, hereafter L-S, attempt to show the conditions under which firms in monopolized industries tend to lessen the durability of their products, compared with the durability of these goods produced under perfect competition. They conclude that when unit costs $c(N)$ of a good are a function of the life N of the good, a monopolist will usually tend to produce goods of a less durable nature. In the case when $c'(N)/c(N)$ is rising, L-S conclude that a monopolist will always produce goods of lower durability, (p. 105). These conclusions are in accord with previous writers who have addressed themselves to the same problem.[1]

The point which is illustrated in this comment is a relatively simple one. Consumers desire durable goods for the services that these goods render over time. The services provided by a good at a point of time do not depend on the life or durability of the good. If the monopolist sells the service directly (rents the product) or indirectly (adopts a sales policy rather than a rental policy), the firm incurs the cost both of producing and replacing the durable good. This result is obvious since, by definition, the monopolist is the only supplier. Therefore, the choice of durability by a monopolist is essentially one of minimizing the cost of the provision of any given service flow from a stock of durable goods. The decision is independent of demand or revenue conditions. It follows that a monopolist will produce goods of the same durability as competitive firms since cost minimization is also achieved under competition.

It is shown that L-S have conceived the problem in such a way that a specification error is introduced. The assumptions on which their analysis is based are, at least in part, contradictory. Their monopoly profit (objective) function based on these assumptions is therefore inappropriate for the analysis of the durability question. One implication of this comment is that earlier writers are also mistaken in their conclusions.[2]

Moreover, in Sections III and IV it is shown that when durability depends not only on the length of life but also on the rate of deterioration (or evaporation) of the product, the optimal degree of durability for the monopoly is still identical with the competitive solution. These more general cases are not considered by L-S. However, R. L. Schmalensee has extended the L-S analysis to include the case in which the product deteriorates at a constant rate. Finally, it is indicated how the argument can be generalized to encompass other types of quality differences between products.

* Monash University, Australia. I would like to thank T. W. Swan for his suggestion that I examine the topic. I would also like to thank Maureen Brunt, David Evans and Ivor Pearce who commented on an earlier draft. All responsibility for errors is my own.

[1] David Martin (p. 227) concludes that a monopolist is likely to restrict the durability of the product when the competitive solution occurs on a major portion of the demand curve. E. Kleiman and T. Ophir (p. 177) conclude that a monopolist will produce "*shorter-lived* assets than would a competitive industry."

[2] Kleiman and Ophir acknowledge the pioneering efforts of Martin. Their analysis of the monopoly problem represents a more general version of the earlier model. L-S appear to be unaware of the earlier literature. The objective (profit) function used by L-S is similar to the one used by Kleiman and Ophir. However, the latter authors consider more than the simple "one-hoss shay" variety of durable good.

I. *Assumptions*

The assumptions on which the analysis of Section II is based are as follows: (Unless otherwise indicated they are the same as those made by L-S.) In the competitive case let there be any number of potential firms each of which can produce output with the same cost function. In the monopoly case, the assumption is that any number of identical plants can be constructed with the same efficiency. Returns to scale in the industry are constant so that choice of technology is independent of output and unit costs are simply a function $c(N)$ of the life of the product where $c'(N)>0$. The product is assumed to provide a constant flow of services until its life N is reached. The product has the property of "sudden death", or is of the "one-hoss shay" variety.

Let p denote the rental rate or the price of a unit of service of the durable good where p is a function

$$p = g(Q) \tag{1}$$

of the total quantity of services available at a point of time.

It is optimal for the monopolist to arrange the production program in such a way that the profit maximizing stock of durable goods is reached as quickly as possible. As a reflection of the optimal program and the assumption of constant returns to scale it is assumed that a quantity Q of the good is produced in the first period. Additional production of Q units takes place at intervals of N periods as replacement production is required. There is no presumption in favor of the L-S assumption of continuous production in each period of y units which would, in time, achieve an even age-distribution of the product.[3]

Although there is no presumption in favor of a constrained production path for the monopolist the effect of the introduction of constraints like those imposed by L-S are interesting. In the Appendix a proof is indicated to show that if there is a constant upper limit to production, w, in each period such that $y(t)\leq w$, where $y(t)$ is current output, the monopolist's choice of durability is unaffected. In the case where the constraint is inoperative and $w\geq Q^*$, where Q^* is the optimal stock, the production path is as illustrated in Figure 1b.

If, in addition, the monopolist is constrained to produce a constant amount in each and every period, i.e., $y(t)=y=Q^*/N$ the choice of durability N affects the speed at which the optimal stock Q^* can be reached. In this constrained optimalization problem demand conditions enter into the choice of N by the monopolist and by competitive firms. However, there are no a priori grounds in favor of the hypothesis that the monopoly will produce less durable goods.

II. *Optimal Life With One-Hoss Shay Obsolesence*

I would suggest the following analysis. If we denote the rate of interest or discount r, then P_N, the capitalized price of the product with N periods of life, will be

$$P_N = p\int_{t=0}^{N} e^{-rt}dt = \frac{p}{r}(1-e^{-rN}) \tag{2}$$

In the case of monopoly control of the industry, the present (or discounted) value

[3] Consider the following quotation from Frederick and Vera Lutz.

There is . . . no reason for presuming that a synchronized process, or stock of durable goods of even age-distribution will be built up. For, so long as the output (or service) stream from the single unit of the good is constant over its lifetime, and so long also as the interest rate is constant for all scales of investment, there exists no mechanism which is bound to bring about an even age-distribution. The present case may be likened to that of electric light bulbs; so long as their performance is independent of their age, the evening out of their age-distribution serves no purpose. [pp. 123–24]

π of all future receipts less payments is given by

$$\pi = [P_N Q - Qc(N)] \cdot [1 + e^{-rN} + e^{-2rN} + \ldots + \ldots] = Q[P_N - c(N)]/[1 - e^{-rN}] \tag{3}$$

since Q goods are produced and sold every N periods. Expression (3) has been simplified by summing the geometric progression. The substitution of (1) and (2) into (3) yield

$$\pi = [Qg(Q)]/r - [Qc(N)]/[1 - e^{-rN}] \tag{4}$$

Now suppose that the monopolist adopts a policy of selling the service rather than directly selling the product. In the case of a rental policy, rental receipts in each period equal $Q\,g(Q)$ since a constant equilibrium stock Q is maintained. The discounted value of these receipts for an infinite planning horizon is

$$Qg(Q) \int_0^\infty e^{-rt}dt = [Qg(Q)]/r \tag{5}$$

Initial (capital) costs and replacement outlays $[Qc(N)]/(1-e^{-rN})$ are the same both in the case of sale and product rental. For a monopoly industry this is obvious given that there is only the one supplier. Therefore, discounted monopoly profits are identical whether the product is rented or sold. This result is as it should be given the (implicit) assumption of a perfect capital market with perfect foresight on the part of both producers and consumers.

The first right-hand (revenue) term in brackets in (4) is independent of the life of the product.[4] It is, therefore, only a formality to show that the choice of durability is the same under monopoly and competition. Expression (4) is differentiated partially with respect to Q and N. The two necessary conditions for profit maximization are found by setting these derivatives to zero:

$$\frac{\partial \pi}{\partial Q} = \frac{Qg'(Q) + g(Q)}{r} - \frac{c(N)}{1 - e^{-rN}} = 0 \tag{6}$$

and

$$\frac{\partial \pi}{\partial N} = \frac{-Q}{(1 - e^{-rN})^2}[(1 - e^{-rN})c'(N) - c(N)re^{-rN}] = 0 \tag{7}$$

The following result is obtained from (6):

$$c(N) = P_N\left(1 - \frac{1}{\eta}\right) \tag{8}$$

where η, $0 < \eta < \infty$, is the elasticity of demand for the services of the durable good. Under competition $c(N) = P_N$ so that the monopoly produces a smaller stock and sells it at a higher than under competition. The simplification of (7) yields:

$$\frac{c'(N)}{c(N)} = \frac{r}{e^{rN} - 1} \tag{9}$$

While it is true that equation (8) is a function of Q and N, equation (9) is a function of N alone. Therefore the solution to (9) gives the optimum product life N selected by the monopoly industry. Since it is independent of (8) it is identical with the competitive industry solution. Incidentally, it agrees with the L-S solution in the competitive case but not with their monopoly result as shown on page 103.

The reason L-S reach the conclusion that the monopoly will produce less durable products is that monopoly profits are implicitly confined to *replacement* sales. When profits derived from initial sales are

[4] This independence does not hold if it is arbitrarily assumed that the firm's planning (or profit) horizon is a variable dependent on the life of the product. For a horizon which is fixed in chronological time or for an infinite horizon monopoly revenue is independent of product lifetime. A similar problem arises in the choice by an entrepreneur of a machine (producer durable) where the entrepreneur's horizon is determined by the life of the machine (see F. and V. Lutz, pp. 120–22). See also the Appendix below.

included, choice of durability is the same under both market structures. Current output y is assumed constant by L-S so that the stock of the durable good builds up gradually: y in period one, $2y$ in period two, ty in period t up to Ny in period N. The stock Ny is stationary since the new additions equals replacement after stocks have built up for N periods. L-S write as follows: "If the durability in a certain stationary situation is N, and the amount of product produced each period is y, then the total services available at any point of time is Ny, . . ." (page 102).

As is shown above, this statement represents a contradiction since in the interval t=0 to N, the quantity of services is rising and the rental price $p=g(ty)$ is falling as time passes.[5] Consider Figure I in which the time paths of production and the rental price are displayed.

Figure 1a depicts the time path of production and the rental price on the L-S assumption of continuous production. On the other hand, Figure 1b shows an intermittent production path which yields a constant flow of services and a constant retail price on the assumptions of the present analysis. L-S ignore the time paths of stock and price until N periods after a particular durability has been chosen.[6]

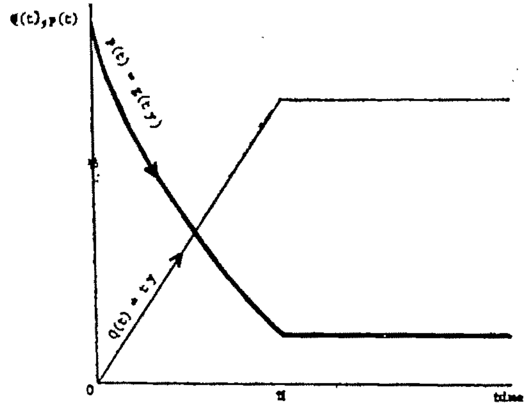


FIGURE 1a

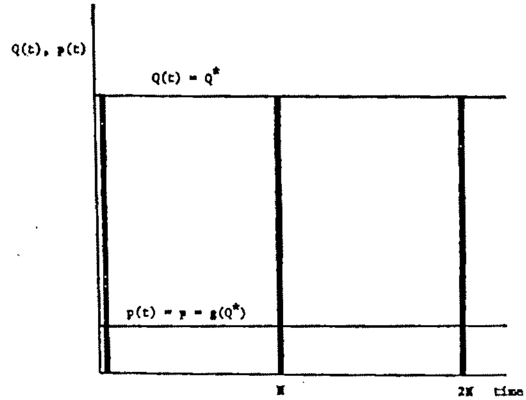


FIGURE 1b

[5] F. and V. Lutz (p. 108–9) point to the highly artificial assumptions under which gross revenue earned by a producer durable machine affects the choice as to the life of the machine:

> This solution, however, rests on a particular assumption about the nature of the "building-up" process; it presupposed, by analogy with the tree case, that the rate of installation of new machines, throughout the building-up process and afterwards, is a constant, and that the output from the total machine stock gradually increases during the building-up process, until, at the moment when the process is completed, it reaches its maximum level, at which it is thereafter maintained. Although this description of the building-up process is appropriate to investment in circulating capital (e.g. the tree case), and has usually been regarded in capital theory as applying equally well to investment in durable goods, as regards the latter it is highly artificial, at least so long as we assume that the demand curve for the product is expected to be constant over time.

The analogy between choice of durability of producer and consumer durable goods is complete. Both involve minimization of the cost of the service flow. The exception to this rule is when output is constrained to be a constant in every period. In this case the need to build stock quickly affects the choice of durability (see the Appendix for amplification).

[6] L-S maximize $yP_N\text{-}yc'(N)$ where y is current output and P_N is defined as the *equilibirum* capitalized price since it is a function of the *equilibrium* capitalized stock Ny (see L-S expression (6) p. 103). The L-S maximand becomes $y[(1-e^{-rN})/r]g(Ny)-yc(N)$. As is argued in the text this maximand refers only to replacement sales and ignores entirely the course of prices and outputs and the need to discount future profits at the time, t=0, a decision to produce the product with a particular life N is made.

Difficulties which arise when monopoly profits are confined to replacement sales are illustrated in the following example. As an extreme case, suppose technology is such that a good has to be produced with an infinite life span or not at all. The good can be produced at low cost and has a high capitalized or rental price. Application of the L-S profit equation would lead to the absurd conclusion that the monopolist should *not* produce any units of the good since replacement sales (y in L-S notation) would be zero despite high discounted returns on the basis of initial sales or as the result of a rental policy.

One implication of the (implicit) assumption by L-S of a perfect capital market is that it is immaterial whether the firm's policy is to sell the product directly or to sell the service (rent the product). It was shown above how the profit function used in the present analysis is the same with either alternative. However, this is not the case in the analysis of L-S since rental receipts take account of initial production whilst replacement sales do not.[7] L-S define sales revenue as

$$y \frac{1 - e^{-rN}}{r} g(Ny) \tag{10}$$

while current revenue derived from a rental policy with an equilibrium stock of Ny units is

$$yNg(Ny) \tag{11}$$

However, expression (10) is not identically equal to (11) because

$$N \neq \frac{1 - e^{-rN}}{r} \tag{12}$$

III. *The Optimal Rate Of Deterioration*

In this section the analysis is extended to the case in which the good produced by the monopolist deteriorates over time. We suppose that in place of the one-hoss shay assumption a durable good is produced which shrinks or evaporates at an exponential rate b where unit costs are a function $c(b)$ of the rate of shrinkage and $c'(b) < 0$. The capitalized price P_b of one unit of the durable good which deteriorates at a rate b and has a rental price p is

$$P_b = p \int_0^\infty e^{-(r+b)}dt = p/(r+b) \tag{13}$$

Let $y(t)$ denote current output of the good and $Q(t)$ the stock of the good in period t. The rate of change of the stock $Q(t)$ with respect to time $dQ(t)/dt$ is given by

$$dQ(t)/dt = y(t) - bQ(t) \tag{14}$$

since a proportion b of the stock evaporates at each point of time. A perfect capital market is assumed. Under these conditions it is immaterial whether the firm has a rental or sales policy. The monopoly industry profit function for a rental policy will be derived since its derivation is easier than for a sales policy in which customers must impute capital gains or losses on their purchases as industry price alters.

The present (or discounted) value of the stream of rental receipts less the discounted cost of producing and maintaining the stock $Q(t)$ of goods is

$$\pi = \int_{t=0}^\infty [Q(t)g(Q(t)) - y(t)c(b)]e^{-rt}dt \tag{15}$$

$$= \int_0^\infty \left\{Q(t)g(Q(t)) - \left[\frac{dQ(t)}{dt} + bQ(t)\right]c(b)\right\} e^{-rt}dt \tag{16}$$

where $Q(0)=0$, $y(0)=0$, $Q(t)$, $y(t) \geq 0$ for all t and $g(Q(t))$ is the rental price of a unit of the good.

The maximization of (16) yields the optimal path of the stock and output over time and the equilibrium value of the

[7] See fn. 5 above.

stock $Q(t)$ when $d\,Q(t)/dt=0$. As well, the optimal value of b, the rate of shrinkage for the monopolized industry, can be derived. These optimal values can be found either by application of the Pontryagin Maximization Principle or by the calculus of variations. The second method is more familiar and will therefore be used here.

The Euler necessary condition for an optimum is

$$\frac{\partial f}{\partial Q(t)} - \frac{d}{dt}\left[\frac{\partial f}{\partial\left(\frac{dQ(t)}{dt}\right)}\right] = 0 \tag{17}$$

where $f[Q(t), d\,Q(t)/dt, t]$ represents the expression inside the integral sign of (16). Evaluation of (16) yields one of the necessary conditions for an optimum:

$$Q(t)g'(Q(t)) + g(Q(t)) = c(b)[r + b] \tag{18}$$

where both revenue and costs have been expressed as current account items. Expression (18) says that the additional current revenue derived from renting the marginal unit of the consumption good must equal the current interest payment on the initial cost of that unit, $r\,c(b)$, plus the current replacement cost of that unit, $b\,c(b)$. Since the left-hand side of (18) is decreasing in $Q(t)$ and the right-hand side is constant the solution must be unique.

Monopoly profit maximization requires that the cost of both producing and maintaining any unit of stock, $(r+b)\,c(b)$ be minimized. The same condition applies to competitive firms. Also the choice of technology is independent of the scale of output or, size of stock. It follows that the choice of b must be the same with both market structures.

A more formal method for the derivation of the optimal degree of durability, b, is to proceed as follows: The terms on the right-hand side of (18) are independent of time. Therefore the stock $Q(t)$ must also be independent of time and the optimal stock,

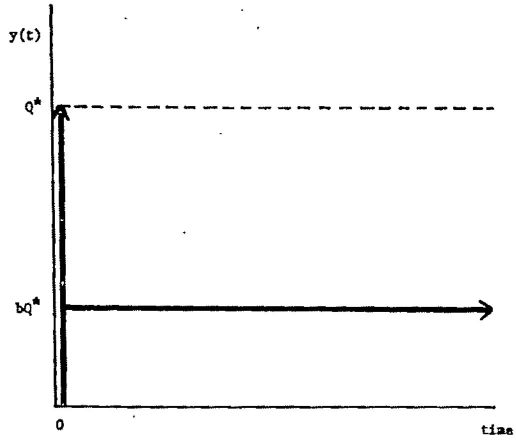


FIGURE 2

call it Q^*, which satisfies (18) must be reached immediately production is commenced.[8] By the nature of the problem the initial stock $Q(0)$ is zero. All subsequent output for $t>1$ is replacement production and is equal to $y(t)=bQ^*$, $(t>1)$. Figure 2 displays the optimal time path of output by means of solid lines. The broken line shows the constant flow of services over time.

The monopoly profit function (16) may now be expressed as follows without the need for time subscripts:

$$\pi = [Q^*g(Q^*) - bQ^*c(b)] \cdot \int_0^\infty e^{-rt}dt - Q^*c(b) \tag{19}$$

$$= \frac{Q^*g(Q^*) - (b+r)Q^*c(b)}{r} \tag{20}$$

The final right-hand term in (19), $Q^*c(b)$, represents the initial cost of reaching the equilibrium stock Q^*. Notice that once again the revenue term in (20), $Q^*\,g(Q^*)/r$, does not depend on the durability term, b.

The other first-order condition for maximization of profits is found by setting the

[8] Helpful assistance with this part of the problem was received from Professor Hirofumi Uzawa at a seminar at Tokyo University, August 1969.

partial derivative of (20) with respect to b equal to zero. This yields

$$(21)\qquad \begin{cases} b + r = \dfrac{-c(b)}{c'(b)} \\ \text{or} \\ b = \left|\dfrac{c(b)}{c'(b)}\right| - r \end{cases}$$

The competitive industry solution can be found from the equality:

$$(22)\qquad c(b) = P_b = p/(r + b)$$

where the rental price p is given for atomistically competitive firms. The derivative of (22) with respect to b yields $c'(b) = -p/[(r+b)^2]$. p is eliminated with the aid of (22) to yield the value of b which is obtained under competition. This value is identical with the monopoly solution, expression (21).

This condition for optimal durability, expression (22), implies that the optimal rate of deterioration or shrinkage of the good is greater the higher is the rate of interest or discount r.[9]

IV. *The Optimal Rate of Deterioration Combined With a Variable Life*

This section illustrates how it is possible to determine simultaneously both the optimal life, N, and the optimal rate of shrinkage, b, when technology is such that there is a trade off between these two components of durability and the initial unit cost, $c(N, b)$. The function $c(N, b)$ is assumed to have the following properties:

$$(23)\qquad \frac{\partial c(N, b)}{\partial N} > 0; \quad \frac{\partial c(N, b)}{\partial b} < 0$$

The capitalized price P_{bN} of one unit of the good is

$$(24)\qquad \begin{aligned} P_{bN} &= p\int_{t=0}^{N} e^{-(b+r)t}dt \\ &= [p(1 - e^{-(b+r)N})]/(r + b) \end{aligned}$$

where p is the rental price. The actual stock Q of the good may be divided into two components by means of the identity

$$(25)\qquad QP_{bN} \equiv \overline{QP}$$

where $\overline{Q}$ is the "effective" stock of the good which has infinite durability and therefore provides a constant stream of services indefinitely. $\overline{P}$ is the capitalized price of an everlasting good of infinite durability. It is equal to

$$(26)\qquad \overline{P} = p\int_{t=0}^{\infty} e^{-rt}dt = p/r$$

Therefore, the actual stock Q may be expressed as

$$(27)\qquad \begin{aligned} Q &= [\overline{PQ}]/P_{bN} \\ &= [(r + b)\overline{Q}]/[r(1 - e^{-(r+b)N})] \end{aligned}$$

with the aid of (24), (25), and (26). In expression (27) a simple transformation of units which does not affect market valuations such as $Q\ P_{bN}$ has been used to partition the actual stock Q of a good with life N and rate of shrinkage b into two components, the effective stock $\overline{Q}$ which is independent of N and b and another term which is dependent on the values of N and b.

The cost of production of an actual stock of Q units is $Q\ c(N, b)$ which equals

$$(28)\qquad Qc(N, b) = [(r + b)c(N, b)\overline{Q}] \,/\, [r(1 - e^{-(r+b)N})]$$

where the substitution is made from (27). Consequently, the cost of production of

[9] The second-order condition for either profit maximization or cost minimization is

$$d^2[(r + b)c(b)]/db^2 > 0$$

which implies that

$$2c'(b) + (r + b)c''(b) > 0$$

But

$$db/dr = -c'(b)/[2c'(b) + (r + b)c''(b)]$$

The denominator is positive by the second-order condition. The numerator is also positive since $c'(b)<0$. Therefore $db/dr>0$ and a rise in the rate of interest results in a less durable product.

one unit of effective stock with infinite durability is found by dividing expression (28) by the effective stock $\bar{Q}$. The resultant expression:

$$[(r+b)c(N,b)]/[r(1-e^{-(r+b)N})] \tag{29}$$

represents the initial (capital) outlay required to produce a good of any desired degree of durability plus the discounted sum of the stream of replacement outlays required to maintain a unit of the good which provides a constant stream of services.

A necessary condition for monopoly profit maximization is the minimization of (29) since when this condition is satisfied total costs (initial outlay plus discounted replacement costs) are at a minimum for any given stock of durable goods. The monopoly incurs these costs directly if it sells the service (rents the product) and indirectly if it sells the product since it is, by definition, the only source of supply.

The partial derivatives of (29) with respect to N and b are set to zero. Two of the necessary conditions for profit maximization and cost minimization are obtained after simplification:

$$e^{(r+b)N} = 1 + [(r+b)c(N,b)] \Big/ \left[\frac{\partial c(N,b)}{\partial N}\right] \tag{30}$$

and

$$\left[\frac{\partial c(N,b)}{\partial b} + \frac{c(N,b)}{r+b} + Nc(N,b)\right] \cdot e^{-(r+b)N} = \frac{\partial c(N,b)}{\partial b} + \frac{c(N,b)}{r+b} \tag{31}$$

The results obtained in Section II are a special case of (30). Let $b=0$ in (30) and expression (9) is obtained. Also, the results obtained in Section III are a special case of (31). Let $N \rightarrow \infty$ in (31) and the left-hand term approaches zero. The result of this operation is that expression (21) is obtained.

The term $e^{(r+b)N}$ may be eliminated from (30) and (31) and the result of this operation is a single equation:

$$c(N,b) = N\frac{\partial c(N,b)}{\partial N} + (r+b)\left|\frac{\partial c(N,b)}{\partial b}\right| \tag{32}$$

Notice that the optimal values of N and b which satisfy (33) are independent of demand conditions for any profit maximizing firm—including a monopoly. Naturally, the values of b and N chosen by competitive firms are the same since survival depends on cost minimization over time.

It can be quite simply argued that in a perfect capital market demand conditions in the product market must be irrelevant for the choice of the optimal degree of durability since units of the same good varying only in durability are identical (perfect substitutes) when an allowance is made for the present value of the future stream of replacement outlays required to maintain any given flow of services (or any given stock).

Moreover, by implication it may be possible to find a numeraire such as the effective stock for some other types of goods subject to quality differences which do not arise from the degree of durability. If prices are identical when expressed in terms of the numeraire then they are perfect substitutes. In these cases also, monopoly and competitive firms should carry out the same process of minimizing the cost of providing any given flow of services rendered by the product. The *quality* of the good produced by the monopolist should be identical with the *quality* chosen by competitive firms.

Suppose further that there is a group of substitute products or that the quality variations of a product are such that they are imperfect substitutes. It can be shown

that under certain conditions a single monopoly will produce the same products and quality variations as a collection of competitive firms subject to the threat of entry (see my article).

Appendix

This Appendix considers the effects of the imposition of constraints on the optimal production path of the monopolist when the consumption good is of the one-hoss-shay variety.

Peter Dixon of Harvard University has kindly offered a proof of the proposition that a production constraint of the form

$$y(t) \leq w \tag{A1}$$

where $y(t)$ is current output and w, an arbitrary constant, does not affect the choice of durability of a consumer good produced by a monopolist. The life N of the product is identical to the competitively determined life. The problem is to maximize

$$\pi = \int_0^\infty [Q(t)g(Q(t)) - c(N)y(t)]e^{-rt}dt \tag{A2}$$

subject to

$$y(t) \leq w,$$
$$dQ(t)/dt = y(t) - y(t-N),$$

and

$$Q(t), \quad y(t) \geq 0$$

The expression in the square brackets of (A2) represents rental receipts less costs incurred in period t.

The Lagrangian expression

$$\begin{aligned}\pi^* = \int_0^\infty \{ & [Q(t)g(Q(t)) - c(N)y(t)]e^{-rt} \\ & - \lambda_1(t)[dQ(t)/dt - y(t) + y(t-N)] \\ & - \lambda_2(t)[y(t) - w]\} dt\end{aligned} \tag{A3}$$

is formed. Two optimality conditions are as follows:

$$\begin{aligned}\frac{\partial \pi^*}{\partial Q(t)} - \frac{d}{dt}\left(\frac{\partial \pi^*}{\partial(dQ(t)/dt)}\right) & = [g(Q(t)) + g'(Q(t))Q(t)]e^{-rt} \\ & + \frac{d\lambda_1(t)}{dt} = 0\end{aligned} \tag{A4}$$

and

$$\frac{\partial \pi^*}{\partial(y(t))} = -c(N)e^{-rt} + \lambda_1(t) - \lambda_1(t+N) - \lambda_2(t) \geqq 0$$

Now $\lambda_2(t) \geqq 0$ where ">" implies $y(t) = w$ so

$$-c(N)e^{-rt} + \lambda_1(t) - \lambda_1(t-N) \geqq 0 \tag{A5}$$

Now

$$\frac{d\lambda_1(t)}{dt} = -MR(t)e^{-rt} \tag{A6}$$

where MR is the expression in square brackets in (A4). This implies

$$\begin{aligned} -\int_T^{T+N}\left[\frac{d\lambda_1(t)}{dt}\right]dt & = \lambda_1(T) \\ - \lambda_1(T+N) & = \int_T^{T+N} MR(t)e^{-rt}dt \end{aligned} \tag{A7}$$

i.e.,

$$\int_T^{T+N} MR(t)e^{-r(t-T)}dt \geqq c(N) \tag{A8}$$

with the aid of (A5). Expression (A8) implies that the present value of the extra revenue from a unit increase in $y(t)$ must be expanded until it equals marginal cost $c(N)$. This condition will hold unless $y(t)$ is already expanded to its limit. The solution is to accumulate at rate w until Q^* is reached where Q^* satisfies

$$MR(Q^*)\int_0^N e^{-rt}dt = c(N)$$

i.e.,

$$MR(Q^*) = \frac{c(N)r}{1 - e^{-rN}} \tag{A9}$$

Since Q^* is independent of N it is clear that

the choice of N does not depend on Q^*, nor on demand conditions. The partial derivative of (A3) with respect to N plus knowledge of the optimal path of output can be used to establish in a rigorous fashion the optimal value of N. The solution is the same as the competitive result, expression (9) above.

Let production be carried out under the same constraint as was imposed by L–S. Namely, current output $y(\mathrm{t})$ is assumed constant in each period. The problem is to maximize

$$\text{(A10)} \quad \pi = \int_0^\infty [Q(\mathrm{t})g(Q(\mathrm{t})) - c(N)y]e^{-r\mathrm{t}}d\mathrm{t}$$

where

$$dQ(\mathrm{t})/d\mathrm{t} = y \qquad \mathrm{t} = 0, \ldots, N$$
$$dQ(\mathrm{t})/d\mathrm{t} = 0 \qquad \text{otherwise.}$$

So

$$\text{(A11)} \quad \begin{aligned} \pi = &\int_0^N [\mathrm{t}yg(\mathrm{t}y) - c(N)y]e^{-r\mathrm{t}}d\mathrm{t} \\ &+ \int_N^\infty [Nyg(Ny) - c(N)y]e^{-r\mathrm{t}}d\mathrm{t} \end{aligned}$$

The optimality conditions are[10]

$$\text{(A12)} \quad \begin{aligned} \frac{\partial\pi}{\partial N} = &[g(Ny) + Nyg'(Ny)]e^{-rN} \\ &- c'(N) = 0 \end{aligned}$$

and

$$\text{(A13)} \quad \begin{aligned} \frac{\partial\pi}{\partial y} = &\int_0^N \mathrm{t}[g(\mathrm{t}y) + \mathrm{t}yg'(\mathrm{t}y)]e^{-r\mathrm{t}}d\mathrm{t} \\ &- \frac{c(N)}{r} \\ &+ \frac{N}{r}e^{-rN}[g(Ny) + g'(Ny)Ny] \end{aligned}$$

The life N is determined by the condition

$$\text{(A14)} \quad M + \frac{Nc'(N) - c(N)}{r} = 0$$

[10] Solutions to this problem were kindly provided by Jean Clare Richard of the Catholic University of Louvain, Belgium, and Peter Dixon.

where M represents the first (integral) term in (A13). For small N, $M \rightarrow 0$ and so

$$\frac{c'(N)}{c(N)} \simeq \frac{1}{N}$$

In the unconstrained maximization problem

$$\frac{c'(N)}{c(N)} = \frac{r}{e^{rN} - 1}$$

from (9). A Taylor's series expansion of e^{rN} shows that in the unconstrained problem $c'(N)/c(N) \simeq 1/N$ also. These results indicate that if N is small and the build-up period $0, \ldots, N$ can be neglected the constrained and unconstrained results are the same.

Suppose a firm in a competitive industry is faced with the problem of the selection of N given the constraint of an unchanged output level y. The maximand or objective function (A10) and (A11) will be the same. However, the competitive firm cannot affect the industry rental price g. The condition required to select the competitive choice of N is

$$\text{(A15)} \quad \begin{aligned} &\int_0^N [\mathrm{t}g(\mathrm{t}y)]e^{-r\mathrm{t}}d\mathrm{t} \\ &+ \frac{N}{r}c'(N) - \frac{c(N)}{r} = 0 \end{aligned}$$

found by differentiating (A11) with respect to N and y for a given g (or industry price).

The only difference between the monopoly and competitive solutions (A14) and (A15), respectively, appears in the first or integral term. The monopoly takes account of marginal revenue during the build-up period while competitive firms are only concerned with average revenue. The choice of durability by the monopoly may be less than, equal to, or greater than the competitive choice depending on the nature of the demand curve for the durable product. A priori there is no reason to predict a bias in one direction or the other.

Neither expression (A14) nor (A15) agree with the respective constrained monopoly and competitive results derived by L–S. In

each case they neglect the integral (or build-up) term and they confine their analysis to replacement sales in the case of the monopoly.

REFERENCES

E. Kleiman and T. Ophir, "The Durability of Durable Goods," *Rev. Econ. Stud.*, Apr. 1966, *33*, 165–78.

D. Levhari and T. N. Srinivasan, "Durability of Consumption Goods: Competition Versus Monopoly," *Amer. Econ. Rev.*, Mar. 1969, *59*, 102–09.

F. Lutz and V. Lutz, *The Theory of Investment of the Firm*, Princeton 1951.

D. D. Martin, "Monopoly Power and the Durability of Durable Goods," *Southern Econ. J.*, Jan. 1962, *28*, 271–77.

L. S. Pontryagin and assoc., *The Mathematical Theory of Optimal Processes*, Oxford 1964.

R. L. Schmalensee, "Regulation and the Durability of Goods," *Bell J. Econ. Manage. Sci.*, 1970, *1*, 54–64.

P. L. Swan, "Market Structure and Technical Progress: The Influence of Monopoly on Product Innovation," *Quart. J. Econ.*, Nov. 1970, *84*.

Agricultural Productivity Differences Among Countries

By Yujiro Hayami and V. W. Ruttan*

The sources of productivity growth over time, and of productivity differences among countries and regions have emerged as a central unifying theme of growth theory and development economics.[1] In recent years a concensus seems to have emerged to the effect that productivity growth in the agricultural sector is essential if agricultural output is to grow at a sufficiently rapid rate to meet the demands for food and raw materials that typically accompany urbanization and industrialization.[2] Failure to achieve rapid growth in agricultural productivity can result either in the drain of foreign exchange or in shifts in the internal terms of trade against industry, and thus seriously impede the growth of industrial production. Failure to achieve rapid growth in labor productivity in agriculture can also raise the cost of transferring labor, and other resources, from the agricultural to the nonagricultural sector as development proceeds.

Extremely wide differences in agricultural productivity exist among countries. Agricultural output per worker in India is approximately one-fiftieth of that in the United States. Relatively few underdeveloped countries have achieved levels of output per worker one-fifth as high as in the United States. Furthermore, these differences have widened during the last decade.[3] This lag in the rate of productivity growth in agriculture represents a serious constraint on economic growth in many developing economies.

Recent empirical research supports a classification of the sources of productivity differences, or of productivity growth, into three broad categories, a) resource endowments, b) technology, as embodied in fixed or working capital, and c) human capital, broadly conceived to include the education, skill, knowledge and capacity embodied in a country's population. Although this is clearly an oversimplification it does represent a substantial advance over the earlier emphasis on a single key or strategic factor.[4]

Our analysis indicates that the three broad categories outlined above account for approximately 95 percent of the differences in labor productivity in agriculture between a representative group of Less Developed Countries (*LDC*'s) and of Developed Countries (*DC*'s). In this comparison the three factors are of roughly equal

* Associate professor, department of economics, Toyko Metropolitan University now visiting associate professor, department of agricultural economics, University of Minnesota, and professor, department of agricultural economics, University of Minnesota, respectively. The authors are indebted to Zvi Griliches, Richard Nelson, Willis Peterson and Mathew Shane for suggestions and comments; and Miss Sachiko Yamoshita and Mrs. Barbara Miller for computational assitance in the preparation of this paper. Minnesota Agricultural Experiment Station Miscellaneous Journal Series Paper No. 1387. The research on which this paper is based was financed through a grant from the Rockefeller Foundation.

[1] J. R. Hicks has suggested that growth theory and development economics have no connection. This view would seem to be invalid in view of Hicks' own criteria. See A. O. Krueger.

[2] See articles by Irma Adelman and Cynthia T. Morris, Dale W. Jorgenson, Gustav Ranis and J. C. H. Fei, and V. W. Ruttan.

[3] See Hayami and associates.

[4] See studies by Zvi Griliches, A. O. Krueger, R. R. Nelson, and T. W. Schultz.

importance. When compared to the *DC*'s of recent settlement (Australia, Canada, New Zealand, and the United States) favorable resource endowments account for somewhat more than one-third of the differences. Resource endowment is the major factor accounting for differences in labor productivity between the *DC*'s of recent settlement and the older *DC*'s. Nevertheless it seems apparent that the *LDC*'s could, over time, achieve labor productivity levels in agriculture well over half as high as in the more recently settled *DC*'s, roughly comparable to the levels achieved in the older *DC*'s, through increased use of technical inputs supplied from the industrial sector and improvements in the quality of the labor force, even in the absence of substantial changes in man-land ratios.

I. *The Method and the Data*

The approach used in this study involves the estimation of a cross-country production function of the Cobb-Douglas type for thirty-eight developed and underdeveloped countries.[5] Differences in agricultural output per worker are accounted for by differences in the level of conventional and nonconventional inputs per worker, classified as a) internal resource accumulation, b) technical inputs supplied by the nonagriculture sector, and c) human capital.[6] All the data used in this study are taken from a recent compilation of international agricultural production statistics by Yujiro Hayami and associates.[7]

Production functions were estimated for three different periods; 1955 (1952–56 averages), 1960 (1957–62 averages), and 1965 (1962–66 averages).[8] The analysis was conducted in gross output (net of seeds and feed) terms in order to include the effects of current intermediate inputs such as fertilizer. Individual agricultural commodities were aggregated by the farm gate (or import) prices of the United States, Japan, and India, to produce three different output series. The series were then averaged geometrically into a single composite output series which was used as the dependent variable.[9]

The independent variables used in the study include labor, land, livestock, fertilizer, machinery, education, and technical manpower. In summing up the effects of resource endowments, technology, and human capital on productivity per worker, land and livestock serve as proxy variables for internal resource accumulation; machinery and fertilizer for technical inputs; and general and technical education in agriculture for human capital.

Land (measured by hectares of agri-

[5] Countries included are: Argentina, Austria, Australia, Belgium, Brazil, Canada, Ceylon, Chile, Colombia, Denmark, Finland, France, Germany, Greece, India, Ireland, Israel, Italy, Japan, Mauritius, Mexico, Netherlands, New Zealand, Norway, Peru, Philippines, South Africa, Spain, Surinam, Sweden, Switzerland, Syria, Taiwan, Turkey, U.A.R., U.K., U.S.A., and Venezuela.

[6] For a report on a preliminary attempt see Hayami (1969, 1970). Major extensions from the previous study include: a) a comprehensive revision of data; b) introduction of the livestock variable; c) analysis on a per farm basis in addition to a national aggregate basis; d) test of stability of the production function over time; and e) refinements in the procedures used to account for productivity differences.

[7] The basic data were collected from publications by the United Nations organizations (*FAO*, *ILO* and *UNESCO*), the Organization for Economic Cooperation and Development, and the governments of various countries. These data were processed by Hayami and associates to be consistent with the definitions of variables and, also, to be comparable among countries. Earlier estimates of agricultural outputs reported by Hayami and Inagi were substantially revised for this study.

[8] Averages were taken for flow variables (output and fertilizer input). Stock variables were in principle measured by 1955, 1960, and 1965 levels. It would seem more consistent to have averages of 1953–57, 1958–62, and 1963–67, but the original estimates of agricultural output are of 1957–62 averages (see Hayami and associates) and, when we tried to extend the 1958–62 output series to 1955 and 1965, the *FAO* index of agricultural production was available only until 1966.

[9] This procedure was applied for 1960 data. 1955 and 1965 output estimates were extrapolated from the 1960 estimates by using the *FAO* indexes of agricultural production by countries.

cultural land) used for agricultural production cannot be regarded as a mere gift of nature. It represents the result of previous investment in land clearing, reclamation, drainage, fencing, and other development measures. Similarly, livestock (as measured by livestock units) represents a form of internal capital accumulation. Thus, in our perspective, land and livestock represent a form of long-term capital formation embodying inputs supplied primarily by the agricultural sector.[10] Both high inputs of land and of livestock per worker tend to be associated with low levels of labor and high levels of land per unit of output. In contrast, fertilizer (as measured by the $N+P_2O_5+K_2O$ in commercial fertilizers) and machinery (as measured by tractor horsepower) represent inputs supplied by the industrial sector. Technical advances stemming from both public and private sector research and development are embodied in or complementary to these modern industrial inputs. Mechanical innovations are usually associated with larger inputs of power and machinery. Biological improvements, such as the innovations embodied in high yielding varieties, are typically associated with higher levels of fertilizer use. In this analysis these two industrial inputs represent proxies for the whole range of inputs which carry modern mechanical and biological technologies.

The proxies for human capital include measures of both the general educational level of the rural population and specialized education in the agricultural sciences and technology. Two alternative measures of the level of general education were attempted: a) the literacy ratio and b) the school enrollment ratio for the primary and secondary levels. Both sets of data are deficient in that they apply to the entire population and are not sensitive to differences in the quality of rural and urban education. Education in the agricultural sciences and technology was measured by the number of graduates per ten thousand farm workers from agricultural faculties at above the secondary level. These graduates represent the major source of technological and scientific personnel for public sector agricultural research and extension and for research development and marketing in the private agribusiness sector.[11]

A critical assumption in this approach is that the technical possibilities available to agricultural producers in the different countries can be described by the same production function. Cross-section production functions, using individual countries or regions as observations, have been widely used. Cross-country aggregate production functions for the agricultural sector were first estimated by Jyoti Bhattacherjee in 1953. An aggregate agricultural production function similar to that used in this study, using states in the United States as observations was employed by Zvi Griliches in an attempt to account for the impact of research and education on agricultural output. Anne Krueger's recent efforts to estimate the contribution of factor endowment differentials to variations in per capita income employs the assumption that all countries are subject to a uniform production function.

In a recent paper Richard Nelson has argued that the assumptions of a common production function ". . . get in the way of understanding international differences in productivity—particularly differences between advanced and underdeveloped countries" (p. 1229). Nelson's objections appear directed primarily to the empirical

[10] Perennial plants belong to the same category of inputs as livestock; but they are not included due to the lack of data.

[11] In a sense this variable may be superior as the proxy for the level of research and extension to the "state average of public expenditure on research and extension per farm" used by Griliches, because our variable reflects the research and extension activities in the private sector as well as in the public sector.

results obtained from use of relatively primitive two-factor production functions, where intercountry differences in value-added per worker are related to the capital-labor ratio. He insists, as a result of differential diffusion of new technology, that "... at any given time one would expect to find considerable variation among firms with respect to the vintage of their technology, certainly between countries, but even within a country" (p. 1230).

We share the Nelson perspective. Agricultural producers in different countries, in different regions of the same country, and on different farms in the same region are not all on the same micro-production function. This reflects differences among producers in their ability to adopt new technology. More importantly, it is also the result of differential diffusion of agricultural technology, and, to an even greater degree, of differential diffusion of the scientific and technical capacity to invent and develop new mechanical, biological, and chemical technology specifically adapted to the factor endowments and prices in a particular country or region.

We may call the envelope of all known and potentially discoverable activities a secular or "meta-production function." The full range of technological alternatives described by the meta-production function is only partially available to individual producers in a particular country or agricultural region during any particular historical "epoch."[12] It is, however, potentially available to agricultural scientists and technicians.

We view the common or cross-country production function which we have estimated as a meta-production function. It is assumed that the invention and diffusion of a new "location specific" agricultural technology through the application of the concepts of physical, biological, and chemical science and of engineering, craft, and husbandry skills, is capable of making the factor productivities implicit in the cross-country production function available to producers in less developed countries. It is also assumed that the capacity of a country to engage in the necessary research, development and extension is measured by the two proxy variables for human capital, namely general education and technical education in agriculture. It appears to us that this effort, and that of Griliches and Krueger, are not inconsistent with the perspective presented by Nelson in his criticism of the empirical results obtained from two factor cross-country production functions.

The production function employed in this study was of the Cobb-Douglas type. It was used mainly because of its ease in manipulation and interpretation. A test presented in the Appendix indicates that the unitary elasticity of substitution implicit in the Cobb-Douglas production function is an acceptable assumption. The ordinary least squares estimation

[12] In the *short run*, in which substitution between capital and labor is circumscribed by the rigidity of existing capital and equipment, production relationships are best described by an activity with relatively fixed factor-factor and factor-product ratios. In the *long run*, in which the constraints exercised by existing capital disappear and are replaced by the fund of available technical knowledge, including all alternative feasible factor-factor and factor-product combinations, production relationships can be adequately described by the neoclassical production function. In the *secular period* of production, in which the constraints given by the available fund of technical knowledge are further relaxed to admit all potentially discoverable knowledge, production relationships can be described by a meta-production function which describes all potentially discoverable technical alternatives. The meta-production function can be regarded as the envelope of neoclassical production functions. Although the term is not employed, the meta-production function concept is implicit in the work of Murray Brown and of W. E. G. Salter. We have discussed the rationale for the meta-production function concept in Japanese and *U.S.* agricultural development in greater detail elsewhere (see Hayami and Ruttan). The elasticity of substitution among factors increases continuously as the time period increases from the *short run* to the *secular* period.

procedure was used. The possibility of simultaneous equation bias seems small because all inputs, except fertilizer, are measured in stock terms and can be treated as predetermined. In a few cases, however, the method of instrumental variables was tried to see if any different inferences might be drawn. The assumption of a common production function among countries is a testable hypothesis. However, it appears that the data used in this study are too crude to be employed for such a test.[13]

II. *Estimation of the Production Function*

We conducted an especially detailed analysis for 1960 because of a) better comparability of output data and b) availability of data for the number of farms in that year.[14] Table 1 presents the estimates of the unrestricted Cobb-Douglas production function on the cross-country data; each column reports the results of a regression of agricultural output on a different set of inputs in the *log linear* form, including estimates of the production elasticities and their standard errors (in parentheses), the standard errors of estimate and the coefficients of determination adjusted for the degrees of freedom.[15] The estimation was made both on per farm data (output and conventional inputs deflated by the number of farms) and on national aggregate data. The results from these two sets of data are not sufficiently different to lead to different inferences regarding the agricultural production structures among countries.

Considering the crudeness of data, the levels of statistical significance of the estimated coefficients seem satisfactory in most cases. The coefficients stay fairly stable when nonconventional variables are added or subtracted, though the coefficients for labor and livestock tend to move opposite to the coefficient for machinery. The results of estimation by the method of instrumental variables (denoted as *IV*) compared with the least square estimates provide no *prima facie* evidence against the use of least squares.

Attempts to include other variables, e.g., the ratio of irrigation land to total land area and the ratio of cropland to pasture land, were tried in an attempt to adjust for differences in the quality of land input; but it turned out that the coefficients for such variables are either negative or nonsignificant.[16]

Plausibility of the estimates may be checked by a comparison with the results of earlier attempts to estimate aggregate production functions in various countries. Bhattacharjee obtained aggregate production elasticities for his cross-country production function (including only conventional variables) centered on 1950 of around 0.3 for labor; 0.3 to 0.4 for land; and 0.3 for fertilizer. The coefficients for livestock and tractors were not significant at commonly accepted levels. The Bhat-

[13] In order to test the assumption that farmers in different countries face the same production function, the production function was estimated separately for the two different groups of countries (*DC* and *LDC*'s). The estimation was tried for various groupings of *DC*'s and *LDC*'s, but the results are all implausible with most of the coefficients statistically nonsignificant or negative in sign. It seems that measurement errors in our observations (especially of nonconventional variables) are too large to make it possible to estimate the influences of variables for the groups of countries within which the ranges of data variations are relatively small. The basic assumption is, therefore, not testable on the presently available data. All we can claim is that differences in agricultural productivity among countries can be explained well with this assumption.

[14] The *1960 World Census of Agriculture* provides the data of the number of farms for a large number of countries. Comparable data are available for only a small number of countries for 1955 and 1965. See also fn. 9.

[15] Surinam was dropped from the sample except Regressions 1 and 6 because of the lack of technical education data.

[16] This does not necessarily mean that such variables have no significant influence, but rather it means that the presently available data are too crude to estimate the influences of such variables.

tacharjee results indicate higher production elasticities for land and fertilizer than the results obtained in our study. It would appear that our model is somewhat better specified in that we obtained statistically meaningful coefficients for livestock and machinery as well as for the two proxy variables for human capital.

The aggregate production elasticities of *U.S.* agriculture were estimated by Griliches as 0.4 to 0.5 for labor; 0.1 to 0.2 for land, fertilizer and machinery; 0.3 to 0.5 for education; 0.04 to 0.1 for research and extension. It is rather surprising that the Griliches' estimates, despite the completely different nature of the data used, coincide so well with the ones in this study.

The production elasticities estimated for Japanese agriculture by Yasuhiko Yuize in value-added terms are in the ranges of 0.4 to 0.6 for labor and 0.2 to 0.4 for land. Such figures are consistent with the estimates in this study since according to the social account study by

TABLE 1—ESTIMATES OF AGRICULTURAL PRODUCTION FUNCTION ON CROSS-COUNTRY DATA, 1960 (1957–62 AVERAGES)

	Per farm Basis						
Regression number	(1)	(2)	(3)	(4)	(5)	(2-*IV*)	(3-*IV*)
Sample size	38	37	37	37	37	37	37
Labor (L)	0.336 (0.121)	0.432 (0.114)	0.393 (0.117)			0.490 (0.110)	0.454 (0.113)
Land	0.071 (0.074)	0.108 (0.065)	0.097 (0.067)	0.117 (0.062)	0.104 (0.066)	0.108 (0.069)	0.097 (0.072)
Livestock	0.166 (0.099)	0.241 (0.089)	0.227 (0.092)	0.249 (0.086)	0.232 (0.091)	0.210 (0.094)	0.192 (0.097)
Fertilizer	0.174 (0.055)	0.124 (0.058)	0.136 (0.062)	0.121 (0.053)	0.126 (0.059)	0.096 (0.058)	0.108 (0.067)
Machinery	0.205 (0.061)	0.057 (0.067)	0.104 (0.064)	0.038 (0.053)	0.092 (0.059)	0.074 (0.068)	0.124 (0.064)
General education Literacy ratio (E_1)		0.348 (0.186)				0.366 (0.196)	
School enrollment ratio (E_2)			0.360 (0.247)				0.263 (0.274)
Technical education		0.190 (0.057)	0.148 (0.055)	0.197 (0.055)	0.146 (0.054)	0.197 (0.060)	0.153 (0.056)
$L\times E_1$				0.418 (0.109)			
$L\times E_2$					0.383 (0.114)		
Coef. of det. (adj.)	0.908	0.932	0.926	0.934	0.928	0.928	0.921
S. E. of est.	0.138	0.119	0.124	0.118	0.123	0.123	0.128
Sum of conventional coefficients	0.952 (0.098)	0.962 (0.085)	0.957 (0.088)	0.943 (0.074)	0.937 (0.080)	0.978 (0.088)	0.975 (0.094)

TABLE 1—(*Continued*)

	National Aggregate basis						
Regression number	(6)	(7)	(8)	(9)	(10)	(7-*IV*)	(8-*IV*)
Sample size	38	37	37	37	37	37	37
Labor (L)	0.335 (0.064)	0.451 (0.074)	0.413 (0.075)			0.474 (0.072)	0.434 (0.074)
Land	0.056 (0.065)	0.088 (0.062)	0.076 (0.063)	0.097 (0.061)	0.080 (0.063)	0.092 (0.065)	0.080 (0.067)
Livestock	0.191 (0.096)	0.247 (0.089)	0.235 (0.092)	0.263 (0.086)	0.243 (0.091)	0.219 (0.093)	0.205 (0.095)
Fertilizer	0.161 (0.053)	0.112 (0.059)	0.123 (0.063)	0.105 (0.058)	0.108 (0.061)	0.090 (0.057)	0.104 (0.064)
Machinery	0.192 (0.056)	0.071 (0.065)	0.116 (0.060)	0.040 (0.053)	0.102 (0.058)	0.082 (0.065)	0.127 (0.061)
General education Literacy ratio (E_1)		0.326 (0.187)				0.321 (0.196)	
School enroll- ment ratio (E_2)			0.324 (0.248)				0.290 (0.271)
Technical education		0.182 (0.057)	0.142 (0.055)	0.195 (0.055)	0.139 (0.054)	0.182 (0.060)	0.142 (0.056)
$L \times E_1$				0.464 (0.072)			
$L \times E_2$					0.432 (0.072)		
Coef. of det. (adj.)	0.955	0.953	0.950	0.954	0.950	0.951	0.948
S. E. of est.	0.131	0.118	0.123	0.118	0.122	0.120	0.125
Sum of conventional coefficients	0.935 (0.035)	0.969 (0.039)	0.963 (0.040)	0.969 (0.039)	0.965 (0.040)	0.957 (0.048)	0.950 (0.040)

Source: Hayami and associates.

Notes: Equations linear in logarithms are estimated by the least squares except those denoted as *IV*, which are estimated by the instrumental variable method. The standard errors of coefficients are in parentheses.

Dependent variable: Gross agricultural output net of seeds and feed in thousand wheat units (one wheat unit is equivalent to a ton of wheat).

Labor: Number of male workers active in agriculture in thousands.

Land: Area of agricultural land in thousand hectares.

Livestock: Livestock in agriculture in thousand livestock units (conversion factors: 1.1 for camels; 1.0 for buffalo, horses and mules; 0.8 for cattle and asses; 0.2 for pigs; 0.1 for sheep and goats; 0.01 for poultry).

Fertilizer: Sum of N, P_2O_5 and K_2O in thousand metric tons contained in commercial fertilizers consumed.

Machinery: Horsepower of tractors for farm purposes in thousands hp.'s.

Literary ratio: Literacy ratio in percent.

School enrollment ratio: Ratio of school enrollments in the primary and secondary schools in percent, adjusted for differences in the school system.

Technical education: Number of graduates from agricultural schools of the third level (college level) per ten thousand male farm workers.

Number of farms: Number of agricultural holdings in thousands.

the Japanese Ministry of Agriculture and Forestry the ratio of value-added to gross output was around 0.7 in Japanese agriculture in the period when Yuize's study was made. In the less developed countries we do not have comparable estimates of the aggregate agricultural production function. Theodore Schultz has, however, inferred from the impact of the 1918–19 influenza epidemic that the production elasticity of labor in Indian agriculture was 0.4. This is consistent with our estimates. Such consistency with other studies gives support to the results of estimation in this study.

Griliches has found that in *U.S.* agriculture, a given percentage increase in education, which improves the quality of labor, has the same output effect as an equal percentage increase in labor itself. In order to test whether the same assertion holds in the international dimension, we have estimated the production function by combining labor L and general education E in a multiplicative form $L \times E$; this resulted in little change (compare regressions 2 with 4, 3 with 5, 7 with 9, and 8 with 10). Furthermore, the analysis of variance provides evidence in support of the equality in the coefficients of labor and general education.[17]

Judging from the sums of coefficients of conventional inputs, compared with the standard errors of those sums (shown in parentheses below the sums of coefficients), constant returns seem to prevail both on the farm firm level and on the national aggregate level. Note, however, that increasing returns prevail when both private and socially controlled inputs are allowed to vary. The constant returns at the farm firm level may explain the existence of farms of extremely different sizes producing the same commodities. The constant returns at the national aggregate level might be one of the distinctive characteristics of agricultural production and, if so, would have important implications for the intersectoral investment priorities for national economic development.

The stability of the agricultural production function over time is tested on the 1955, 1960, and 1965 cross-country sample. Because comparable data on the number of farms were not available for 1955 and 1965, we assumed the linear homogeneity in the Cobb-Douglas production function and regressed output per capita (per male worker) on conventional inputs per capita and on nonconventional inputs. The linear homogeneity assumption is based on the information contained in Table 1. In order to make the data comparable among years we restricted the countries included in the sample to 36 (Mauritius and Surinam were dropped from the sample for lack of labor data).

The results of our estimations are summarized in Table 2. Comparing the estimates of the per capita production function with those of the unrestricted form in Table 1, we see that the land coefficients become smaller and the livestock coefficients become larger. This appears to be caused by high intercorrelation between land area per worker and livestock per worker. Differences in the two sets of estimates do not seem to imply different conclusions. The production parameters seem largely stable over time. The null hypothesis of the equality of the production coefficients among 1955, 1960, and 1965 is accepted according to the results of analysis of variance (the F-statistic calculated from Regressions 12, 13, 14, and 17 is only 0.95).

III. *Accounting for Productivity Differences*

The results obtained from estimation of the agricultural production function in

[17] The F-statistics calculated for testing the equality of the labor and education coefficients are: 0.22 for Regression 2 vs. Regression 4; 0.31 for Regression 3 vs. Regression 5; 0.65 for Regression 7 vs. Regression 9; 0.77 for Regression 8 vs. Regression 10.

TABLE 2—ESTIMATES OF AGRICULTURAL PRODUCTION FUNCTION ON CROSS-COUNTRY DATA: 1955 (1952–56 AVERAGES); 1960 (1957–62 AVERAGES); 1965 (1962–66 AVERAGES)

	Per-capita basis						
Regression number	(11)	(12)	(13)	(14)	(15)	(16)	(17)
Year	1960	1960	1955	1965	1955–60	1960–65	1955–60–65
Sample size	36	36	36	36	72	72	108
Land	0.072 (0.061)	0.056 (0.063)	0.082 (0.061)	0.043 (0.073)	0.068 (0.042)	0.047 (0.047)	0.066 (0.038)
Livestock	0.289 (0.092)	0.281 (0.094)	0.311 (0.093)	0.273 (0.101)	0.300 (0.064)	0.276 (0.066)	0.286 (0.055)
Fertilizer	0.105 (0.057)	0.107 (0.063)	0.124 (0.057)	0.142 (0.083)	0.120 (0.041)	0.125 (0.049)	0.137 (0.038)
Machinery	0.076 (0.063)	0.125 (0.059)	0.061 (0.049)	0.152 (0.063)	0.090 (0.036)	0.144 (0.041)	0.106 (0.032)
General education Literacy ratio	0.362 (0.180)						
School enroll- ment ratio		0.337 (0.243)	0.168 (0.182)	0.356 (0.336)	0.320 (0.141)	0.324 (0.189)	0.243 (0.134)
Technical education	0.182 (0.055)	0.137 (0.053)	0.194 (0.051)	0.099 (0.050)	0.168 (0.035)	0.113 (0.034)	0.122 (0.029)
Dummy: 1960					−0.009 (0.026)		−0.017 (0.029)
1965						−0.019 (0.029)	−0.021 (0.030)
Coef. of det. (adj.) S. E. of est.	0.934 0.115	0.930 0.119	0.931 0.111	0.919 0.135	0.934 0.111	0.929 0.123	0.924 0.122
Implicit coefficient of labor	0.458	0.431	0.422	0.390	0.422	0.408	0.405

Notes and Source: See Table 1.

Equations linear in logarithms are estimated by the least squares. The standard errors of coefficients are in parentheses.

the previous sections may be used to account for intercountry differences in labor productivity (output per male worker) in agriculture in 1960.

Since our production function is now assumed to be linear homogeneous (with respect to conventional inputs) in the Cobb-Douglas form, the percentage difference in output per worker can be expressed as the sum of percentage differences in conventional inputs and nonconventional inputs per worker each weighted by the relevant production elasticities. Based on the results shown in Table 1 and Table 2 the following set of production elasticities was adopted: 0.40 for labor, 0.10 for land, 0.25 for livestock, 0.15 for fertilizer, 0.10 for machinery, 0.40 for education, and 0.15 for research and extension. Only the school enrollment ratio

TABLE 3—ACCOUNTING FOR DIFFERENCE IN LABOR PRODUCTIVITY IN AGRICULTURE BETWEEN DEVELOPED COUNTRIES (*DC*) AND LESS DEVELOPED COUNTRIES (*LDC*) AS PERCENT OF THE LABOR PRODUCTIVITY OF *DC*

	Group 1 (13 *DC's*)		Group 2 (9 *DC's*)		Group 3 (4 *DC's*)	
Difference in output per male worker—percent	88.8	(100)*	83.5	(100)	93.6	(100)
Percent of difference explained: Total	84.2	(95)	71.1	(85)	90.0	(96)
Resource accumulation:	29.2	(33)	17.5	(21)	32.6	(35)
Land	9.2	(10)	1.8	(2)	9.7	(10)
Livestock	20.0	(23)	15.7	(19)	22.9	(25)
Technical inputs:	24.3	(27)	24.3	(29)	24.5	(26)
Fertilizer	14.5	(16)	14.5	(17)	14.6	(16)
Machinery	9.8	(11)	9.8	(12)	9.9	(10)
Human capital:	30.7	(35)	29.4	(35)	32.9	(35)
General education	18.2	(21)	17.6	(21)	19.5	(21)
Technical education	12.5	(14)	11.7	(14)	13.4	(14)

* Inside of parentheses are percentages with output per worker set equal to 100.
LDC: Brazil, Ceylon, Colombia, India, Mexico, Peru, Philippines, Syria, Taiwan, Turkey, UAR.
DC: Australia, Belgium, Canada, Denmark, France, Germany, Netherlands, New Zealand, Norway, Sweden, Switzerland, UK, USA.
Group 1 includes all *DC*'s;
Group 2 excludes Australia, Canada, New Zealand, and the United States from *DC*'s;
Group 3 includes only the four *DC*'s excluded from Group 2.
Accounting formula:

$$\left(\frac{Y_d - Y_l}{Y_d}\right) = 0.10\left(\frac{a_d - a_l}{a_d}\right) + 0.25\left(\frac{s_d - s_l}{s_d}\right) + 0.15\left(\frac{f_d - f_l}{f_d}\right) + 0.10\left(\frac{m_d - m_l}{m_d}\right) + 0.40\left(\frac{E_d - E_l}{E_d}\right) + 0.15\left(\frac{U_d - U_l}{U_d}\right)$$

where y, a, s, f, m, are, respectively, output, land, livestock, fertilizer, machinery per male worker; E and U are, respectively, the general education (school enrollment ratio) and the technical education variable; lower case letter d denotes *DC* and l denotes *LDC*.

was used as the education variable in this accounting, but the results would have been essentially the same if the literacy ratio had been used.

Two alternative sets of results are presented. The first set involves group comparisons between *LDC*'s and *DC*'s. The second set involves individual comparisons of selected *LDC*'s and *DC*'s with the United States.

Group Comparisons

The sources of differences in labor productivity between the eleven *LDC*'s and different groups of *DC*'s are presented in Table 3. Each column compares for each group the percentage difference in agricultural output per worker between *LDC*'s and *DC*'s with the percentage differences in input variables weighted by the specified production elasticities. Inside of the parentheses is shown the index with the output-per-worker difference set equal to 100. The countries classified as *LDC*'s, for the purposes of this comparison, all had per capita income of less than 350 *U.S.* dollars and more than 35 percent of their labor force engaged in agriculture. The countries classified as *DC*'s had per capita income higher than 700 *U.S.* dollars and less than 30 percent of the labor force engaged in agriculture. Countries falling between these criteria are not included in the comparisons presented in Table 3.

The difference in average agricultural output per worker between the eleven *LDC*'s and the thirteen *DC*'s of group 1

was 88.8 percent; the difference between the eleven *DC*'s and the nine older *DC*'s of group 2 was 83.5 percent; and the difference between the eleven *LDC*'s and the four *DC*'s of recent settlement—group 3—was 93.6 percent. The six variables included in the production function accounted for 95, 85, and 96 percent of the difference in agricultural output per worker between the *LDC*'s and the three *DC*'s groups.

In the comparison between the eleven *LDC*'s and the thirteen *DC*'s—group 1—each generalized category, internal resource accumulation (land and livestock), technical inputs from the industrial sector (fertilizer and machinery), and human capital (general and technical education in agriculture), account for approximately one-third of the explained difference in labor productivity.

The main difference between group 1 and the other two groups is the amount of the difference explained by land. Difference in land accounts for only 2 percent of the difference in labor productivity between the *LDC*'s and the older *DC*'s, while it accounts for 19 percent between the *LDC*'s and the new *DC*'s. This implies that it should be feasible for the *LDC*'s, even with the present land-labor ratios to achieve levels of productivity per worker roughly equivalent to the labor productivity levels achieved by workers in the older *DC*'s—that is, roughly four times as high as present *LDC* levels and well over half the level achieved by the *DC*'s of recent settlement. The critical elements in achieving such increases in labor productivity are the supply of modern industrial inputs in which the new technology is embodied and the investment in general education and in research and extension which raises the capacity to develop and adopt a more productivity technology.

Comparison of group 2 and 3 results does indicate that resource endowments, particularly land, do represent a serious barrier to efforts of both that *LDC*'s and the older *DC*'s to achieve levels of output per worker comparable to the levels currently enjoyed in the more recently settled *DC*'s. This is the first time, to our knowledge, that the economic advantage of the favorable resource endowments in these countries has been demonstrated quantitatively.

Individual Comparisons

The individual country comparisons presented in Table 4 were developed in order to provide somewhat deeper insight into the sources of differences in labor productivity between different "ideal type" *DC*'s and *LDC*'s and the United States. Each now compares the percentage difference in agricultural output per worker between each country and the United States with the linear combinations of percentage differences in input variables weighted by the specified production elasticities. Inside of the parentheses is the index with the output-per-worker difference set equal to 100. In general, the results are consistent with the group comparisons.

In the four underdeveloped countries—India, Philippines, United Arab Republic, and Colombia—internal resource accumulation accounts for approximately one-third and technical inputs roughly one-fourth of the differences. Human capital accounts for more than one-third of the difference between the United States and India, the United Arab Republic, and Colombia. In the Philippines, which has achieved a relatively high level of schooling and produces a relatively large number of agricultural college graduates, human capital explains less than one-fourth of the productivity difference. The contrast between India and the Philippines in this respect is quite striking.

In the comparisons between the countries of Europe and the United States, differences in internal resource accumula-

TABLE 4—ACCOUNTING FOR LABOR PRODUCTIVITY DIFFERENCES FROM THE UNITED STATES AS PERCENT OF *U.S.* LABOR PRODUCTIVITY, 11 SELECTED COUNTRIES

	Difference in output per worker from *U.S.* as percent of *U.S.*	Percentage of difference explained by: Total	Resource accumulation (land and livestock)	Technical inputs (fertilizer and machinery)	Human capital (general and technical education)
LDC					
India	97.8	102.1	32.7	25.0	44.4
	(100)[a]	(104)	(33)	(26)	(45)
Philippines	96.2	82.1	33.4	24.9	23.8
	(100)	(85)	(34)	(26)	(25)
UAR	95.6	97.0	33.8	24.6	38.6
	(100)	(101)	(35)	(26)	(40)
Colombia	89.7	89.4	25.8	24.7	38.9
	(100)	(100)	(29)	(28)	(43)
Europe					
Denmark	52.3	51.0	20.4	13.2	17.4
	(100)	(97)	(39)	(25)	(33)
Netherlands	56.6	51.7	25.0	15.0	11.7
	(100)	(91)	(44)	(26)	(21)
United Kingdom	55.8	50.2	18.2	13.4	18.6
	(100)	(90)	(33)	(24)	(33)
France	63.9	64.3	26.2	16.5	21.6
	(100)	(101)	(41)	(26)	(34)
Japan	89.2	66.0	34.1	22.4	9.5
	(100)	(74)	(38)	(25)	(11)
Pastoral farming					
Argentina	60.0	45.9	−4.8	24.3	26.4
	(100)	(76)	(−8)	(40)	(44)
New Zealand	−42.4	−49.1	−55.2	2.7	3.4
	(100)	(116)	(130)	(−6)	(−8)

[a] Inside of parentheses are percentages with output per worker differences set equal to 100.

tion represent the most significant source of difference in labor productivity. The constraint of land on agricultural productivity is relatively modest for the United Kingdom which experienced the drastic agricultural transformation after the repeal of the Corn Law; it is strongest for France which preserved peasant farms by protective tariffs. Increases in the use of technical inputs and improvements in the quality of human capital can bring labor productivity of the several European countries closer to the *U.S.* level. Nevertheless it seems apparent that major advances in labor productivity in European agriculture (especially in countries like France) toward the *U.S.* level are dependent on the absorption of a higher percentage of the agricultural labor force into the nonagricultural sector. The Japanese case is similar to the European, except that Japan, characterized by a stronger constraint of land, has moved further toward the exhaustion of productivity differentials associated with investment in education and research. In our judgment the model underestimates the significance of the land constraint in the Japanese case and, to a lesser degree, in the European case. Without a significant increase in land area per worker it would be impossible for Japanese agriculture to increase technical inputs (especially machinery) to the *U.S.* level.

The two pastoral farming cases are of particular interest. In spite of low levels of technical inputs, labor productivity in Argentina is roughly comparable to that in Europe. This is due almost entirely to a favorable man-land ratio comparable to that in the United States. Argentina has, as a result of under-investment in technology and human capital, failed to fully exploit its favorable man-land ratio. New Zealand, in contrast, has achieved a level of labor productivity well above the *U.S.* level (the highest in the world) by complementing its favorable resource endowments with high levels of technical inputs and investment in education and research.

The results obtained in both group and individual comparisons are somewhat different than those obtained by Krueger. Using a different methodology, Krueger found that human capital explained more than half the difference in income levels between the United States and a group of less developed countries. This is in contrast to our studies in which human capital explains approximately one-third of the difference in labor productivity. Krueger's results apply to the entire economy and ours to only the agricultural sector. It seems reasonable to expect that resource endowments would be of relatively greater significance in the agricultural sector than in the total economy. We see, therefore, no inconsistency between our results and those obtained by Krueger. In general the consistency between the results presented in Tables 3 and 4, combined with our general knowledge of the economies being studied, strengthens our confidence in the methodology employed in this study.

IV. *Implications for Agricultural Development Strategy*

The implications of this analysis for agricultural development strategy in the less developed countries have both encouraging and discouraging aspects. It is clear that output per worker in the several *LDC*'s can be increased by several multiples, while land area per worker remains constant or even declines slightly. To achieve increases of this magnitude will require substantial investment a) in rural education and b) in the physical, biological, and social sciences. The latter is required for the technical and institutional infrastructure needed for the invention, development, and extension of a more efficient agricultural technology. It will also require the allocation of substantial resources to the production of the technical inputs supplied by the industrial sector, by which new technology is carried into agriculture. By and large, these changes achieve the higher levels of output per worker through increases in output per unit area.

A more discouraging aspect of this analysis is that in order to achieve levels of labor productivity comparable to the levels achieved in the *DC*'s of recent origin it will be necessary to complement those technical changes designed to increase output per unit area with technologies that reduce the labor input per unit area. Significant reduction in labor input per unit area is likely to occur, however, only in those economies in which urban-industrial development is sufficiently advanced to absorb not only the growth in the rural labor force but also to permit a continuous reduction in employment in rural areas.[18] It should be noted that this has occurred in Japan only since World War II. In most *LDC*'s it seems likely that the agricultural labor force will continue to expand more rapidly than the non-agricultural demand for labor from rural areas.

The implications for agricultural development strategy for most less de-

[18] See the article by Folke Dovring.

veloped countries seem relatively clear. An attempt must be made to close the gap in the level of modern industrial inputs and in education and research. Agricultural surpluses generated by closing the gap, over and above the amount necessary to maintain the growth of agricultural productivity, must be used to finance industrial development.[19]

Maintenance of the rate of growth of agricultural productivity can be expected to impose a substantial drain on the savings that can be generated from the agricultural surpluses. Initially a substantial component of industrial capacity must be designed to provide technical inputs for the agricultural sector. Substantial investment will be needed to create the institutional infrastructure to improve general education in rural areas and to produce the technical and scientific manpower needed to bring about technical changes in agriculture. Investment in land development, such as irrigation and drainage, will also be necessary in a number of countries in order to obtain a full return from the new biological and chemical technology.

If successful, the effort would, over time, result in a rate of growth in the nonagricultural labor force sufficient to permit a reduction in the agricultural labor force and a rise in labor productivity toward the levels of the *DC*'s of recent settlement. Clearly the process outlined here is inconsistent with the low cost route to agricultural development that seemed to be opened up by the dual economy models which have dominated much of the theoretical discussion of agricultural development during the last decade.

[19] Shigeru Ishikawa has suggested that achievement of national agricultural output and productivity objectives may, in some developing countries, require a net flow of savings from the nonagricultural to the agricultural sector. The possibility has been such a shock to some students of development economics that they recommend a "development without agriculture" policy (e.g., M. J. Flanders).

Appendix

A Test of Unitary Elasticity of Substitution

In the analysis in the text the production function was specified as being of the Cobb-Douglas type, thus assuming unitary elasticity of substitution among inputs. Here we attempt to test this assumption by estimating the parameters of the *CES* production function developed by Kenneth Arrow, Hollis Chenery, Bagicha Minhas and Robert Solow.

The basic models used for estimation are

$$\log (Y/L) = a + b \log W + c \log Z$$

and

$$\log (V/L) = a' + b' \log W + c' \log Z$$

where Y and V are, respectively, gross output and value-added in argiculture; L is labor; W is the wage rate (measured by output); Z is the shorthand notation for nonconventional variables which shift the production function (general and technical education, in the case of this study). It is well known that under competitive factor markets b or b' measures the elasticity of substitution (between labor and the aggregate of other conventional inputs, including current inputs, in the case of b; or between labor and capital in the case of b'). Also, c or c' measures $k(1-b)$ where k is the exponential coefficient of Z, if Z is specified as a multiplicative shifter of the original *CES* production function (see articles by Arrow et al., and by Hayami 1970).

Therefore, in order to accept the hypothesis of unitary elasticity of substitution, a) the estimated parameters of b and b' should not be significantly different from one, and b) the estimated parameters of c and c' should not be significantly different from zero.

The results of estimation from the available data of 22 countries for 1960 (1957–62 averages) are summarized in Table A. This analysis was conducted exclusively by the 1960 set of data because the value-added series were not available for 1955 and 1965. Availability of wage data limited the sample

TABLE A—ESTIMATES OF THE ELASTICITY OF SUBSTITUTION FUNCTION ON CROSS-COUNTRY DATA, 1957–62 AVERAGES

Regression number	Dependent variable	Coefficients of: Wage: Current (1957–62 av.)	Wage: Lagged (1952–56 av.)	General Education: Literacy ratio	General Education: Sch. enrol. ratio	Technical education	Coef. of det.	S.E. of estimate
(A1)	Y/L	1.152 (0.094)					0.878	0.175
(A2)	Y/L		1.112 (0.145)				0.736	0.258
(A3)	Y/L	1.101 (0.159)		0.131 (0.331)			0.872	0.179
(A4)	Y/L	1.106 (0.151)			0.162 (0.408)		0.872	0.179
(A5)	Y/L	0.927 (0.196)		0.155 (0.322)		0.124 (0.085)	0.879	0.174
(A6)	Y/L	0.962 (0.180)			0.107 (0.400)	0.119 (0.086)	0.878	0.175
(A7)	V/L	1.047 (0.098)					0.864	0.171
(A8)	V/L		1.002 (0.149)				0.709	0.250
(A9)	V/L	1.039 (0.165)		0.018 (0.331)			0.855	0.176
(A10)	V/L	1.039 (0.160)			0.024 (0.411)		0.855	0.176
(A11)	V/L	0.886 (0.209)		0.050 (0.328)		0.102 (0.087)	0.858	0.174
(A12)	V/L	0.908 (0.194)		−0.006 (0.407)		0.101 (0.087)	0.858	0.175

Source: Hayami and associates.

Notes: Equations linear in logarithms are estimated by the least squares. The standard errors of coefficients are in parentheses.

Countries included in the sample are 22: Austria, Belgium, Canada, Ceylon, Denmark, Finland, France, Germany, India, Ireland, Japan, Mauritius, Mexico, New Zealand, Norway, Peru, Philippines, Portugal, Sweden, Turkey, United Kingdom and the United States. Finland, Norway, and Sweden are not included in the sample for the estimation of regressions A7-A12.

Definitions of variables are the same as in Table 1 except: V: Value-added in agriculture in thousand wheat units; Wage: Farm wage rate per day per male worker including board in *U.S.* dollars, converted from native currency by official exchange rates.

size to 22. Finland, Norway, and Sweden were discarded from the sample in the value-added term analysis because of the implausible estimates of value-added for these three countries (see Hayami and associates). Two sets of wage data were tried for estimation:

current wage rate (W_t: 1957–62 averages) and lagged wage rate (W_{t-1}: 1952–56 averages).

The lagged wage rate was tried to determine whether the adjustment might not be instantaneous. The results are quite similar, however, because there is a high correlation between current wage and lagged wage. The Koyck-Nerlove type of distributed lag model was also tried. The results were implausible, however, probably because of high intercorrelation between the wage rate and the lagged dependent variables.

Both in gross output terms and in value-added terms the results of estimation are consistent with the unitary elasticity of substitution hypothesis: a) the coefficients of wage rate are not significantly different from one and b) the coefficients of shift variables, general and technical education, are not significantly different from zero at conventional significance levels. There is little evidence against the use of the Cobb-Douglas production function for the cross-country analysis of agricultural production. Such a conclusion seems consistent with the results derived from the cross-regional analysis of agricultural production in the United States by Griliches and in Japan by Hiromitsu Kaneda, although their results are less conclusive with some of the estimates of b being significantly different from one and some of the estimates of c significantly different from zero.

REFERENCES

I. Adelman and C. T. Morris, "An Econometric Model of Socio-Economic and Political Change in Underdeveloped Countries," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1184–218.

K. J. Arrow, H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

J. P. Bhattacharjee, "Resource Use and Productivity in World Agriculture," *J. Farm Econ.*, Feb. 1955, *37*, 57–71.

M. Brown, *On the Theory and Measurement of Technological Change*, Cambridge 1966, 43–62.

F. Dovring, "The Share of Agriculture in a Growing Population," *Mon. Bull. Agr. Econ. Statist.*, (FAO) Aug.-Sept. 1959, *8*, 1–11.

M. J. Flanders, "Agriculture Versus Industry in Development Policy: The Planner's Dilemma Re-examined," *J. Develop. Stud.*, Apr. 1968, *5*, 171–89.

Z. Griliches, "Research Expenditures, Education, and the Aggregate Agricultural Production Function," *Amer. Econ. Rev.*, Dec. 1964, *54*, 961–74.

Y. Hayami, "Sources of Agricultural Productivity Gap Among Selected Countries," *Amer. J. Agr. Econ.*, Aug. 1969, *51*, 564–75.

———, "On the Use of the Cobb-Douglas Production Function in the Cross-country Analysis of Agricultural Production," *Amer. J. Agr. Econ.*, forthcoming.

——— and K. Inagi, "International Comparison of Agricultural Productivities," *Farm Economist*, 1969, *11* (No. 10), 1–13.

——— and V. W. Ruttan, "Factor Prices and Technical Changes in Agricultural Development: The United States and Japan, 1880–1960," *J. Polit. Econ.*, Nov. 1970, *74*.

———, in association with B. Miller, W. Wade and S. Yamashita, *An International Comparison of Agricultural Production and Productivities*, Univ. Minn. Agr. Exp. Sta. Tech. Bull., No. 277, St. Paul, 1970.

J. R. Hicks, *Capital and Growth*, Oxford 1965.

S. Ishikawa, *Economic Development in Asian Perspective*, Tokyo 1968.

D. W. Jorgenson, "The Development of a Dual Economy," *Econ. J.*, June 1961, *71*, 309–34.

H. Kaneda, "Substitution of Labor and Nonlabor Inputs and Technical Change in Japanese Agriculture," *Rev. Econ. Statist.*, Dec. 1968, *47*, 163–71.

A. O. Krueger, "Factor Endowments and Per Capita Income Differences among Countries," *Econ. J.*, Sept. 1968, *78*, 641–59.

R. R. Nelson, "A Diffusion Model of International Productivity Differences in Manufacturing Industry," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1219–48.

G. Ranis and J. C. H. Fei, "A Theory of Economic Development," *Amer. Econ. Rev.*, Sept. 1961, *51*, 533–64.

V. W. Ruttan, "Growth Stage Theories, Dual Economy Models, and Agricultural Development Policy," Publication No. AE 1968/2, Univ. Guelph 1968.

W. E. G. Salter, *Productivity and Technical Change,* Cambridge 1960.

T. W. Schultz, *Transforming Traditional Agriculture,* New Haven 1964.

Y. Yuize, "Nogyo ni okeru Kyoshiteki Seisankansu no Keisoku" (The Aggregate Production Function in Agriculture), *Nogyo Sogo Kenkyu,* Oct. 1964, *18,* 1–54.

Japan, Ministry of Agriculture and Forestry, *Nogyo oyobi Nokano Shakai Kanjo 1967,* (Social Accounts of Agriculture and Farm Households) Tokyo 1968.

United Nations, *1960 World Census of Agriculture,* FAO, New York.

The Neoclassical Theory of Technical Progress

By Winston W. Chang*

Since technology has often been regarded in the neoclassical theory of technical change as a shift parameter of the production function, the investigation of the way in which technical progress affects this function has been of major concern. The customary analysis involves an examination of the effect of progress on the relationship between certain primal and dual variables which are derived from the production function and which include the input-output coefficients, the factor proportions, the marginal productivities, and the marginal rate of substitution. In particular, when a certain relationship is postulated to be invariant, one obtains a specific type of technical neutrality.

The most important types of neutrality are identified with the names of Hicks, Harrod, and Solow. Technical progress is called Hicks-neutral when the relationship between the marginal rate of substitution and the factor proportion is invariant (see J. R. Hicks, p. 121). This type of neutrality has been heavily used in the theory of international trade. When the relationship between the capital-output ratio and the rate of profit remains unaltered, technical progress is Harrod-neutral (see R. F. Harrod, p. 23).[1] This neutrality is necessary for the feasibility of steady growth. The mirror image of Harrod-neutrality is Solow-neutrality[2] in which the relationship between output per man and the wage rate is unchanged under technical progress. The production functions in this case exhibit purely capital-augmenting progress, and were first used by Robert Solow (1962) to estimate embodied technical change. This type of neutrality is also called Fei-Ranis neutrality (E. S. Phelps) for they first introduced the concept of technical changes which are "U-neutral" in a study of underdeveloped economies (see J. C. Fei and G. Ranis 1965).

Just as there is more than one way to classify technical change, there is also more than one way to measure the extent and bias of progress. The commonly used parameters which define both the extent and bias are changes in marginal products or changes in input-output or factor-augmenting coefficients.[3] Yet there are still a variety of parameters remaining to be defined. The first objective of this paper is to present some new basic sets of parameters associated with Harrod's and Solow's classifications in a two-sector economy producing both consumption goods and capital goods. These parameters are the changes, which result from technical ad-

* Associate professor of economics at the State University of New York at Buffalo. I am indebted to Ronald W. Jones for helpful discussions, and to James Gapinski, Ryuzo Sato, and a referee for helpful comments on an earlier draft of this paper. I am also indebted to Edwin Burmeister and Rodney Dobell for their kind permission to cite some of the results in chapter 5 of their forthcoming book.

[1] For the identification of each type of neutrality as the presence of an invariant relationship between variables, see the article by Martin Beckmann and Ryuzo Sato (1969). According to Harrod's original definition (p. 23 and pp. 26–27) technical progress is neutral if it leaves the capital-output ratio unchanged at a given rate of profit.

[2] See, e.g., Frank Hahn and R. C. O. Matthews (p. 830), Sato and Beckmann (1968, p. 59), (1969), E. Burmeister and R. Dobell, etc.

[3] For the discussions on the conditions for factor-augmenting progress, see, e.g., Phelps, H. Rose, Burmeister and Dobell.

vance, in the primal variables (input-output coefficients) with a certain dual variable being held constant. These basic sets of parameters are developed and related to the usual Hicksian rate and bias of technical change. It turns out that there are a number of new ways beyond those commonly used in the literature to express the extent and bias of technical change.

Although there have been extensive discussions on the classifications and occasionally the equivalence of neutralities in the Hicksian and Harrodian schema,[4] it seems that there has not been a unified treatment to clarify the basic relationships existing among Hicks', Harrod's, and Solow's classifications in a two-sector model. The second objective of the present paper is to provide such a unified treatment. The basic relations between any two of the three measures of bias including the cases of "own" neutrality are developed. In particular, the asymmetrical results out of Harrod's and Solow's classifications in relation to that of Hicks are presented. Some fundamental results derived from combining together all three measures of bias are obtained and interpreted. These results help characterize the basic structure of the general equilibrium model and include, as special cases, the results so far obtained in one-sector models.[5]

The third objective is to discuss the conditions for aggregate neutrality. In particular, it is shown that the usual symmetrical results obtained for aggregate neutrality between the Hicksian and Harrodian schema,[6] in general, do not apply to Solow-neutrality. It is the different requirement for the value of the elasticity of substitution between commodities on the demand side that marks the basic asymmetry. Throughout the paper it is assumed that the production functions are linear homogeneous and that inputs are paid according to competitive imputations.

I. *Some Basic Parameters and Preliminary Results*

Consider the general neoclassical production relations involving disembodied technical change, t, capital, K_j, and labor, L_j, in the production of output, Y_j, in industry j.

$$Y_j = F^j(K_j, L_j, t) \qquad j = 1, 2 \tag{1}$$

Let industry 1 be the capital-goods sector and industry 2, the consumption-goods sector. Denoting the relative change in a variable by a hat (^) over that variable, and totally differentiating (1) yields (2), where θ_{ij} is the distributive share of factor i in industry j and where π_j refers to the rate of technical progress in industry j.[7]

$$\hat{Y}_j = \theta_{Kj}\hat{K}_j + \theta_{Lj}\hat{L}_j + \pi_j \tag{2}$$

Define $e_{Kj} \equiv \hat{Y}_j - \hat{K}_j$ and $e_{Lj} \equiv \hat{Y}_j - \hat{L}_j$. The e_{ij} represents the relative change in the productivity of factor i in industry j, or equivalently, the proportional rate of *reduction* in the corresponding input-output coefficient. Rearranging (2) yields (3) and (4) where $k_j \equiv K_j/L_j$.

$$\begin{aligned} e_{Kj} &= \pi_j - \theta_{Lj}\hat{k}_j \\ e_{Lj} &= \pi_j + \theta_{Kj}\hat{k}_j \end{aligned} \tag{3}$$

[4] See, e.g., Akihiro Amano (1964a, 1964b, 1967), A. Asimakopulos and John Weldon, Asimakopulos, Peter Diamond, Fei and Ranis (1963, 1965), C. E. Ferguson, Hahn and Matthews, R. W. Jones (1965a, 1965b, 1966), Kennedy (1961, 1962b), Meade, Phelps, Joan Robinson (1951), Akira Takayama, Hirofumi Uzawa, etc.

[5] In a one-sector model, Fei and Ranis (1965) have examined the basic relationships among Hicks, Harrod, and Solow biases and have clarified the fundamental structure of an aggregate economy. However, there is no straightforward extension to a multi-sectoral economy. Even a two-sector model usually involves changes in the ratio of commodity prices as a result of technical change. Thus the input-output ratios in terms of value units and the marginal products are usually affected by changes in commodity prices.

[6] See Jones (1965a, 1966).

[7] Throughout the paper whenever the subscript i appears, it always denotes K and L; and j always denotes 1 and 2.

Thus the rate of technical change is a share-weighted average of the e_{ij}.

$$\pi_j = \theta_{Kj}e_{Kj} + \theta_{Lj}e_{Lj} \tag{4}$$

Let w be the real wage rate and r the rental per unit of capital, both measured in consumption goods. In a competitive economy with constant returns to scale, sectoral factor proportions depend only on the ratio of factor prices and the state of technology:

$$k_j = f^j(w/r, t) \tag{5}$$

Total differentiation of (5), noting the relation $\hat{k}_j|_{\hat{w}-\hat{r}} = e_{Lj}|_{\hat{w}-\hat{r}} - e_{Kj}|_{\hat{w}-\hat{r}}$, yields

$$\hat{k}_j = \sigma_j(\hat{w} - \hat{r}) + \beta_j, \tag{6}$$

where σ_j is the elasticity of factor substitution in sector j and where β_j is a Hicksian measure of bias conveniently defined in terms of the Hicksian parameters $b_{ij}(\equiv e_{ij}|_{\hat{w}-\hat{r}})$[8]

$$\beta_j = b_{Lj} - b_{Kj} \tag{7}$$

Technological progress is labor-saving, neutral, or capital-saving in Hicks' sense according as β_j is greater than, equal to, or less than zero. Combine (3) and (6) to obtain (8).

$$\begin{aligned} e_{Kj} &= \pi_j - \theta_{Lj}[\sigma_j(\hat{w} - \hat{r}) + \beta_j] \\ e_{Lj} &= \pi_j + \theta_{Kj}[\sigma_j(\hat{w} - \hat{r}) + \beta_j] \end{aligned} \tag{8}$$

Setting the ratio of factor prices constant in both equations in (8), one immediately obtains the relationship expressing each Hicksian parameter, b_{ij}, in terms of the Hicksian rate and bias of technical change:

$$\begin{aligned} b_{Kj} &= \pi_j - \theta_{Lj}\beta_j \\ b_{Lj} &= \pi_j + \theta_{Kj}\beta_j \end{aligned} \tag{9}$$

It follows that π_j is a share-weighted average of the Hicksian parameters as shown in (10).[9]

$$\pi_j = \theta_{Kj}b_{Kj} + \theta_{Lj}b_{Lj} \tag{10}$$

A dual set of parameters (m_{ij}), may now be defined to represent the proportionate rates of change in the marginal productivities at constant factor inputs, resulting from technical change. Euler's theorem and the assumption of constant returns to scale imply (11),

$$Y_j = K_jF_1^j + L_jF_2^j, \tag{11}$$

where $F_1^j \equiv \partial F^j/\partial K_j$ and $F_2^j \equiv \partial F^j/\partial L_j$. Logarithmic partial differentiation of (11), holding inputs constant, yields the familiar expression for the extent of technical progress:

$$\pi_j = \theta_{Kj}m_{Kj} + \theta_{Lj}m_{Lj} \tag{12}$$

Competitive equilibrium insures that the wage/rent ratio is linked to the sectoral factor proportions and the state of technology:

$$w/r = h^j(k_j, t) \tag{13}$$

Differentiating (13) logarithmically, we obtain (14)

$$\hat{w} - \hat{r} = \hat{k}_j/\sigma_j + m_{Lj} - m_{Kj}, \tag{14}$$

which, combined with (6), yields (15).[10]

$$\beta_j = \sigma_j(m_{Kj} - m_{Lj}) \tag{15}$$

If technological progress is always factor-augmenting, we may write the production functions as follows:

[8] The measure of bias employed here is in fact in the spirit of W. E. G. Salter (ch. 3). His measure of bias is the proportionate rate of increase of the least-cost capital-labor ratio for fixed output and factor prices. Nonetheless, the b_{ij} which define β_j will be called the Hicksian parameters for they conform to the Hicksian concept of technical change (see Hicks). The measure of bias (β_j) has been used before by Amano (1964b), Jones (1965a, 1966), and Takayama.

[9] The result in equation (10) has been derived before by Jones (1965a, 1966) and Amano (1967), although the methods of derivation are somewhat different.

[10] Each m_{ij} again can be expressed in terms of π_j and β_j; namely, $m_{Kj} = \pi_j + (1/\sigma_j)\theta_{Lj}\beta_j$ and $m_{Lj} = \pi_j - (1/\sigma_j)\theta_{Kj}\beta_j$. A particular measure of the Hicks bias, $m_{Kj} - m_{Lj}$, has been used, e.g., by Amano (1964a), Diamond, Drandakis and Phelps, Fei and Ranis (1963, 1965), Ferguson, and Meade. Since this particular measure has the same algebraic sign as β_j, both can serve the same purpose. For the present paper, the measure, β_j, is the more convenient one.

$$Y_j = G^j[B_j(t)K_j,\ A_j(t)L_j]$$

The following relations can be derived:[11]

$$(16)\quad \begin{aligned} m_{Kj} &= \hat{B}_j + (\theta_{Lj}/\sigma_j)(\hat{A}_j - \hat{B}_j) \\ m_{Lj} &= \hat{A}_j - (\theta_{Kj}/\sigma_j)(\hat{A}_j - \hat{B}_j) \end{aligned}$$

The extent and bias of technical progress become:

$$(17)\qquad \pi_j = \theta_{Kj}\hat{B}_j + \theta_{Lj}A_j$$

and

$$(18)\qquad \beta_j = (1 - \sigma_j)(A_j - \hat{B}_j)$$

The relations showing the impact of technological progress upon commodity and factor prices are crucial for the analyses in later sections. Making use of the competitive equilibrium conditions $r = F_1^2 = pF_1^1$ and $w = F_2^2 = pF_2^1$, where p is the price of capital goods in terms of consumption goods, and differentiating equation (11) with substitution from (2) yields:

$$(19)\quad \begin{aligned} \pi_1 &= \theta_{L1}(\hat{w} - \hat{p}) + \theta_{K1}(\hat{r} - \hat{p}) \\ \pi_2 &= \theta_{L2}\hat{w} + \theta_{K2}\hat{r} \end{aligned}$$

Since r/p and w/p are, respectively, the rate of profit and the wage rate measured in capital goods, the above relations suggest the well-known dual aspect that the nature of technical progress can be characterized by the way in which the factor-price frontier shifts.[12]

It would be convenient to define π_j/θ_{ij} as T_{ij} and obtain from (19) the following useful and symmetric relations:

$$(20)\qquad \hat{w} - \hat{r} = -(1/\theta_{L1})(\hat{r} - \hat{p}) + T_{L1}$$

$$(21)\qquad \hat{w} - \hat{r} = (1/\theta_{K1})(\hat{w} - \hat{p}) - T_{K1}$$

$$(22)\qquad \hat{w} - \hat{r} = -(1/\theta_{L2})\hat{r} + T_{L2}$$

$$(23)\qquad \hat{w} - \hat{r} = (1/\theta_{K2})\hat{w} - T_{K2}$$

[11] Drandakis and Phelps (p. 831) have developed the relations in (16) for the one-sector case.

[12] For the concept of the factor-price frontier see Paul Samuelson. The fact that technical change shifts the factor-price frontier has been pointed out, e.g., by Solow (p. 55), Hahn and Matthews (p. 830).

There is more than one way to interpret the T_{ij}. For instance, we have

$$(24)\quad \begin{aligned} T_{L1} &= (\hat{w} - \hat{p})\big|_{\hat{r}-\hat{p}=0} = (\hat{w} - \hat{r})\big|_{\hat{r}-\hat{p}=0} \\ T_{L2} &= \hat{w}\big|_{\hat{r}=0} = (\hat{w} - \hat{r})\big|_{\hat{r}=0} \end{aligned}$$

Thus each T_{Lj} represents the percentage shift of the factor-price frontier (or the shift in labor's marginal product or in the wage/rent ratio) at an unchanged marginal product of capital in industry j. Another useful interpretation of the T_{ij} is seen from equation (4). For example, T_{Lj} is the percentage shift in labor's productivity (or efficiency) at an unchanged capital-output ratio; T_{Kj} likewise is the percentage shift in capital's productivity (or efficiency) at an unchanged labor-output ratio.[13] In terms of the unit product isoquant, T_{ij} measures the percentage inward shift horizontally or vertically toward the i-th axis. Clearly, the above two types of interpretations are dual to each other.

If technical progress is purely labor-augmenting, i.e.,

$$Y_j = G^j[K_j,\ A_j(t)L_j],$$

then from (17) we have $T_{Lj} = \hat{A}_j$. Similarly, if progress is purely capital-augmenting, $T_{Kj} = \hat{B}_j$. The T_{ij} will be given a further interpretation in the next section where Harrod's and Solow's classifications of technical change are analyzed.

II. *Some New Parameters; Harrod and Solow Neutrality and "Own" Neutrality*

With respect to the concept of neutrality in Harrod's sense, one examines the direction of change in the capital-output ratio at a constant rate of profit (r/p). Define the new set of parameters q_{ij} as $e_{ij}\big|_{\hat{r}-\hat{p}=0}$ which may be termed the Harrodian parameters (derived from physical

[13] This interpretation has been made before by Amano (1964b) and Jones (1965b, 1966).

units). By substituting (20) into (8), we obtain

$$(25)\qquad \begin{aligned} q_{Kj} &= \theta_{Lj}(T_{Lj} - \sigma_j T_{L1} - \beta_j) \\ q_{Lj} &= \theta_{Kj}(T_{Kj} + \sigma_j T_{L1} + \beta_j) \end{aligned}$$

Thus we have a new measure of Hicksian rate and bias in terms of the Harrodian parameters:[14]

$$(26)\qquad \pi_j = \theta_{Kj} q_{Kj} + \theta_{Lj} q_{Lj}$$

and

$$(27)\qquad \beta_j = q_{Lj} - q_{Kj} - \sigma_j T_{L1}$$

T_{Lj} can now be given an additional interpretation. Since $T_{Lj} = q_{Lj} | q_{Kj=0}$, it is the increase in output per man at an unchanged capital-output ratio and rate of profit. We may therefore conveniently call it the Harrodian rate of technical progress in sector j.[15]

For the capital-goods sector, the parameter q_{K1} can be used as a measure of bias in the Harrodian sense. However, for the consumption-goods sector, the capital-output ratio should be measured in "value" terms. Denoting the proportionate rate of reduction of the capital-output ratio in value terms in the consumption-goods sector by e^{v}_{K2}, we have

$$(28)\qquad e^{v}_{K2} = e_{K2} - \hat{p}$$

From equations (19), we obtain

$$(29)\qquad \hat{p} = (\theta_{L1} - \theta_{L2})(\hat{w} - \hat{r}) + \pi_2 - \pi_1$$

Combining (29), (20), and (22) yields the expressions for the relative changes in the commodity prices in terms of the differential Harrodian rates of technical change $(T_{L2} - T_{L1})$ and the rates of change of factor prices:

$$(30)\qquad \begin{aligned} \hat{p} = &-(1 - \theta_{L2}/\theta_{L1})(\hat{r} - \hat{p}) \\ &+ \theta_{L2}(T_{L2} - T_{L1}) \end{aligned}$$

and

$$(31)\qquad \hat{p} = (1 - \theta_{L1}/\theta_{L2})\hat{r} + \theta_{L1}(T_{L2} - T_{L1})$$

Without technical progress all relative shares, θ_{ij}, can be expressed in terms of various types of elasticities between any two of the commodity and factor prices (p, w, r, w/p, and r/p). (Some of these elasticities are of particular importance in the theory of international trade.) Let $\eta^{z}_{x} = (\hat{z}/\hat{x})\,|_{\pi_1=\pi_2=0}$. For the purposes at hand, we single out η^{p}_{r} which is:[16]

$$(32)\qquad \eta^{p}_{r} = 1 - \theta_{L1}/\theta_{L2}$$

Combining (30), (8), (20), (32), and (28) yields

$$(33)\qquad \begin{aligned} e^{v}_{K2} = &(\theta_{L2}/\theta_{L1})(\sigma_2 - \eta^{p}_{r})(\hat{r} - \hat{p}) \\ &+ \theta_{L2}[(1 - \sigma_2)T_{L1} - \beta_2] \end{aligned}$$

Thus an increase in the rate of profit as of constant technology may not increase the value of the capital-output ratio in the consumption-goods sector unless $\sigma_2 > \eta^{p}_{r}$. The common sense of this condition can perhaps best be revealed by considering the meanings of these two elasticities. Since σ_2 represents the percentage changes in (physical) output-capital ratio resulting from a one percent change in r as of constant technology,[17] the direction of changes in the value of the capital-output ratio naturally depends on the degree of re-

[14] The Harrodian and Hicksian parameters have the following relations:

$$q_{Kj} = b_{Kj} - \theta_{Lj}\sigma_j T_{L1} \quad \text{and} \quad q_{Lj} = b_{Lj} + \theta_{Kj}\sigma_j T_{L1}$$

[15] See also Jones (1966).

[16] The various elasticities have the following relationships:

$$\eta^{p}_{w/r} = -\theta_{L1}\eta^{p}_{r/p} = \theta_{K2}\eta^{p}_{w} = \theta_{K1}\eta^{p}_{w/p};$$
$$\theta_{K1}\eta^{p}_{w/p} + \theta_{L1}\eta^{p}_{r/p} = 0, \qquad \theta_{K2}\eta^{p}_{w} + \theta_{L2}\eta^{p}_{r} = 0;$$
$$\eta^{w/r}_{r/p} = -1/\theta_{L1}, \qquad \eta^{w/r}_{w} = 1/\theta_{K2}, \qquad \eta^{w/r}_{w/p} = 1/\theta_{K1},$$
$$\eta^{w/r}_{r} = -1/\theta_{L2}; \text{ and } \eta^{r}_{w} = -\theta_{L2}/\theta_{K2},$$
$$\eta^{r/p}_{w/p} = -\theta_{L1}/\theta_{K1}$$

The last two are the elasticities of the factor-price frontiers. Obivously, the following identities hold:

$$1/\eta^{p}_{w/r} = 1/\eta^{p}_{w/p} - 1/\eta^{p}_{r/p} = 1/\eta^{p}_{w} - 1/\eta^{p}_{r}; \text{ and}$$
$$\eta^{p}_{w/p} = \eta^{p}_{r/p}\eta^{r/p}_{w/p} \text{ and } \eta^{p}_{w} = \eta^{p}_{r}\eta^{r}_{w}$$

[17] This can be checked by substituting (22) into (8) to obtain $\sigma_2 = (1/\hat{r})e_{K_2}|_{\pi_2=0}$. The symmetrical case where the elasticity of substitution also links changes in output per man to changes in the real wage was discussed in Arrow, Chenery, Minhas, and Solow (p. 229).

sponse of the physical output-capital ratio and the commodity price ratio.[18] Thus the relative magnitudes of the two elasticities play an important role. The sign of η_r^p depends solely on the capital-intensity condition. It is positive only if the consumption-goods sector is less capital-intensive than the capital-goods sector. Thus the rate of profit and the value of the capital-output ratio may move in the opposite directions only if the so-called capital-intensity condition is not satisfied.

Since $q_{K1}=e_{K1}^{*}|_{\hat{r}-\hat{p}=0}$, one may choose the parameter $e_{Kj}^{*}|_{\hat{r}-\hat{p}=0}$ as a measure of Harrod's bias in sector j. For convenience, however, we choose to take $(-1/\theta_{Lj})e_{Kj}^{*}|_{\hat{r}-\hat{p}=0}$ as our measure of Harrods' bias. Denote this by α_j. In view of (25) and (33), we have

$$(34)\qquad \alpha_j = \beta_j - (1-\sigma_j)T_{L1}$$

Technological change is Harrod labor-saving in sector j if $\alpha_j>0$, Harrod neutral if $\alpha_j=0$, and Harrod capital-saving if $\alpha_j<0$.[19] When the elasticity of substitution is unity, it is well known that Hicks neutral is equivalent to Harrod neutral and that the production function is of the Cobb-Douglas type.

It is important to consider some special cases. If technological progress is purely labor-augmenting and the Harrodian rates of progress are equal as between sectors, one may write

$$(35)\qquad \begin{aligned} Y_1 &= G^1[K_1, A(t)L_1] \\ Y_2 &= G^2[K_2, A(t)L_2] \end{aligned}$$

In this case equation (18) reduces to $\beta_j=(1-\sigma_j)\hat{A}$. Since $T_{Lj}=\hat{A}$, we have from (34) that $\alpha_1=\alpha_2=0$; namely, both sectors exhibit Harrod-neutral technical change. It is readily seen from equation (30) that in steady-state equilibrium in which the rate of profit is constant ($\hat{r}-\hat{p}=0$), the ratio of commodity prices remains unchanged ($\hat{p}=0$). As Burmeister and Dobell (1970) have shown, the production possibility frontier in this case can be expressed as[20]

$$(36)\qquad Y_2 = \Phi[Y_1; K, A(t)L],$$

where $K=K_1+K_2$ and $L=L_1+L_2$.

Another special case is the one in which there is no technical change in the capital-goods sector. It follows from (34) that the classifications of technical change according to Harrod's and Hicks' criteria in the consumption-goods sector come to the same thing. In particular, when the rate of profit is constant, Hicks-neutral technological change in the consumption-goods sector implies that the capital-output ratio in value terms in that sector remains unaltered. This is the Kennedy result (1962a, 1962b).[21] Formally, in this Kennedy case, we have

$$(37)\qquad \begin{aligned} Y_1 &= G^1(K_1, L_1) \\ Y_2 &= A(t)G^2(K_2, L_2) \end{aligned}$$

Clearly, $T_{L1}=0$ and $T_{L2}=\hat{A}/\theta_{L2}$. Equation (30) shows that in steady-state equilibrium in which the rate of profit is constant, the price of capital goods in terms of consumption goods rises at the percentage rate $\hat{A}$, i.e., $\hat{p}=\hat{A}$. This is a central result, and it explains why the Kennedy case and the previous Harrod case generate different asymptotic values of the economic variables. The production possibility frontier, as shown in Burmeister and Dobell (1970), becomes[22]

[18] From fn. 8, one can easily derive $\eta_{r/p}=\theta_{L2}/\theta_{L1}>0$. Thus an increase in the rate of profit (r/p) always raises the rental on capital (r) as of constant technology.

[19] Amano (1964b) has used this measure of bias. Diamond has used, in our notation, α_j/σ_j as his measure of bias. Fei and Ranis (1965) in their one-sector model have employed, in our notation, $-q_K$ as their measure of bias. Here the sectoral subscript is, of course, suppressed.

[20] The mirror-image case in which $Y_1=G^1[B(t)K_1, L_1]$ and $Y_2=G^2[B(t)K_2, L_2]$ implies $Y_2=\Phi[Y_1; B(t)K, L]$.

[21] Also see Joan Robinson (1962, pp. 88–91). This reference is due to the referee.

[22] By the stability assumption that the overall capital-labor ratio in physical terms approaches a certain asymptotic value, Burmeister and Dobell (1970) have

$$(38) \qquad Y_2 = A(t)\Phi(Y_1; K, L)$$

One may consider the direction of change in the physical capital-output ratio in the consumption-goods sector at an unchanged "own" rate of return to capital in that sector (cf. Jones (1965b). If $e_{K2}|_{\hat{r}=0} = 0$, own neutrality in physical terms in sector 2 obtains. This type of Harrod's own neutrality is analogous to Harrod neutrality in the capital-goods sector.

I shall analyze another type of Harrod *own* neutrality—one which is expressed in value terms. Substitute (31) and (34) into (33) to obtain:

$$(39) \qquad \begin{aligned} \overset{v}{e}_{K2} &= (\sigma_2 - \overset{p}{\eta}_r)\hat{r} \\ &\quad - \theta_{L2}[(\sigma_2 - \overset{p}{\eta}_r)(T_{L2} - T_{L1}) + \alpha_2] \end{aligned}$$

Define $\alpha_2^0 \equiv (-1/\theta_{L2})\overset{v}{e}_{K2}|_{\hat{r}=0}$ and $\alpha_1^0 \equiv (-1/\theta_{L1})\overset{v}{e}_{K1}|_{\hat{r}-\hat{p}=0}$ as the measures of Harrod's *own* bias in value terms for the two sectors. Clearly $\alpha_1^0 = \alpha_1$. We therefore obtain the general expression:

$$(40) \qquad \overset{0}{\alpha}_j = \alpha_j + (\sigma_2 - \overset{p}{\eta}_r)(T_{Lj} - T_{L1})$$

Harrod's *own* neutrality in value terms in sector 2 is equivalent to Harrod neutrality only if either $\sigma_2 = \eta_r^p$ or $T_{L2} = T_{L1}$. The former condition is necessary for the value of the capital-output ratio in sector 2 to be independent of the rate of profit or the *own* rate of return to capital in that sector as of constant technology; the latter is necessary for the ratio of commodity prices to remain unaffected in the presence of technical progress at an unchanged rate of profit (r/p) or rental on capital goods (r). Note that even if $\pi_1 = \pi_2$, Harrod neutral and *own* neutral in general are not equivalent. Clearly, it is the Harrod rate of technical change, T_{Lj}, not the Hicks rate, π_j, that is the key parameter in the analysis. Finally, substituting (32) and (34) into (40) yields the relation between Harrod and Hicks *own* biases:[23]

$$(41) \qquad \begin{aligned} \overset{0}{\alpha}_j &= \beta_j - (1 - \sigma_j)T_{Lj} \\ &\quad + (\theta_{L1}/\theta_{L2})(T_{Lj} - T_{L1}) \end{aligned}$$

Consider now the classification of technical progress in Solow's schema. In a one-sector model, it is exactly the mirror-image of Harrod's classification with K and L reversed in every case. It is natural to extend the concept of neutrality to the two-sector model and to use the real wage rate in terms of consumption goods as the invariant dual variable.

Substitute (23) into (8) and define the Solow parameters $u_{ij} \equiv e_{ij}|_{\hat{w}=0}$ to obtain:

$$(42) \qquad \begin{aligned} u_{Kj} &= \pi_j + \theta_{Lj}(\sigma_j T_{K2} - \beta_j) \\ u_{Lj} &= \pi_j - \theta_{Kj}(\sigma_j T_{K2} - \beta_j) \end{aligned}$$

Thus the new expressions for the Hicks rate and bias of technical change are:[24]

$$(43) \qquad \pi_j = \theta_{Kj}u_{Kj} + \theta_{Lj}u_{Lj}$$

and

$$(44) \qquad \beta_j = u_{Lj} - u_{Kj} + \sigma_j T_{K2}$$

For convenience, we choose $\gamma_j \equiv u_{Lj}/\theta_{Kj}$ as our measure of bias in Solow's sense.[25] A positive γ_j, for instance, represents Solow labor-saving technical change. By (42) we obtain the relation between Solow and Hicks measures of bias:

$$(45) \qquad \gamma_j = \beta_j + (1 - \sigma_j)T_{K2} + (T_{Kj} - T_{K2})$$

For the consumption-goods sector, Solow

used the production possibility frontier to prove that p is rising in steady-state equilibrium in the Kennedy case.

[23] Hicks neutral is always equal to Hicks' own neutral in each sector for the wage-rent ratio is always equal to the marginal productivity ratio in each sector.

[24] The following relations among Solow's, Hicks', and Harrod's parameters hold:

$$\begin{aligned} u_{Kj} &= b_{Kj} + \theta_{Lj}\sigma_j T_{K2} \\ &= q_{Kj} + \theta_{Lj}\sigma_j(T_{L1} + T_{K2}) \\ u_{Lj} &= b_{Lj} - \theta_{Kj}\sigma_j T_{K2} \\ &= q_{Lj} - \theta_{Kj}\sigma_j(T_{L1} + T_{K2}) \end{aligned}$$

$$u_{Kj} \gtreqless b_{Kj} \gtreqless q_{Kj} \quad \text{and} \quad u_{Lj} \lesseqgtr b_{Lj} \lesseqgtr q_{Lj}$$

[25] Of course, one may choose u_{Lj} as a measure. Fei and Ranis (1965) in their one-sector model use, in our notation, $-u_L$ as their measure of bias.

neutral is equivalent to Hicks neutral only if $\sigma_2 = 1$. For the capital-goods sector, however, the results are quite different. Even if $\sigma_1 = 1$, the two may not be equivalent: it depends on what I shall call the "Solow" rate of progress, T_{Kj}, which represents the proportional rate of reduction in the capital-input coefficient at a constant output per man and real wage rate. Note that the differential Solow rate of progress $(T_{K1} - T_{K2})$ is the sole determinant of the relative commodity price movements at an unchanged real wage rate.[26] If $T_{K1} = T_{K2}$, the ratio of commodity prices would remain unaltered at a constant wage rate, and the relations between the two measures of bias would be analogous in each sector. If technical progress occurs in sector 2 (consumer goods), Hicks and Solow neutrality would be equivalent in sector 1 only if $\sigma_1 = T_{K1}/T_{K2}$. On the other hand, if there is no progress in the consumption-goods sector, Hicks neutrality in the capital-goods sector must obtain Solow labor-saving improvement. This result is not symmetrical to Kennedy's result (1962a, 1962b) concerning the equivalence of Hicks and Harrod neutrality in the consumption-goods sector when there is no progress in the other sector.

To analyze the relationship between Harrod and Solow measures of bias, we eliminate β_j from (34) and (45) to obtain:

$$\begin{aligned}\gamma_j &= \alpha_j + (1 - \sigma_j)(T_{L1} + T_{K2}) \\ &\quad + (T_{Kj} - T_{K2})\end{aligned} \tag{46}$$

Two interesting points emerge. First, for the consumption-goods sector, Harrod and Solow neutral are equivalent if, and only if, $\sigma_2 = 1$. Second, $\sigma_1 = 1$ is neither necessary nor sufficient for the two neutralities to coexist in the capital-goods sector. The two neutralities imply each other only if $\sigma_1 = 1 + (T_{K1} - T_{K2})/(T_{L1} + T_{K2})$; therefore, if the Solow rate of progress is higher (lower) in the capital-goods sector than in the other sector, only if $\sigma_1 > 1 (< 1)$ will the two neutralities be equivalent.[27]

Consider now the case where output per man in sector j remains unchanged at a constant marginal productivity of labor in that sector. This is Solow's own neutrality. Define $u_{i1}^0 \equiv e_{i1}\big|_{\hat{w} - \hat{p} = 0}$ and $u_{i2} = e_{i2}\big|_{\hat{w} = 0}$. Clearly, $u_{i2}^0 = u_{i2}$. By substituting (21) into (8) and making use of (42), we obtain:[28]

$$\begin{aligned}\overset{0}{u}_{Kj} &= \pi_j + \theta_{Lj}(\sigma_j T_{Kj} - \beta_j) \\ \overset{0}{u}_{Lj} &= \pi_j - \theta_{Kj}(\sigma_j T_{Kj} - \beta_j)\end{aligned} \tag{47}$$

I shall choose $\gamma_j^0 \equiv u_{Lj}^0/\theta_{Kj}$ as a measure of Solow's own bias. Hence

$$\overset{0}{\gamma}_j = \beta_j + (1 - \sigma_j) T_{Kj} \tag{48}$$

As expected, Hicks neutral (or *own* neutral) is equivalent to Solow *own* neutral in sector j only if $\sigma_j = 1$.

Before leaving this section, let us examine the relationship between Harrod *own* neutrality and Solow *own* neutrality. By eliminating β_j from equations (41) and (48), we find:

$$\begin{aligned}\overset{0}{\gamma}_j &= \overset{0}{\alpha}_j + (1/\theta_{Lj})(1 - \sigma_j) T_{Kj} \\ &\quad - (\theta_{L1}/\theta_{L2})(T_{Lj} - T_{L1})\end{aligned} \tag{49}$$

For the capital-goods sector, the two *own* neutralities are equivalent only if $\sigma_1 = 1$. This, however, is not necessarily the case in the consumption-goods sector. Even if $\sigma_2 = 1$, the two neutralities will not be present simultaneously unless the Harrodian rates of progress, T_{Lj}, are equal as between sectors. In general, the two neutralities coexist only if $\sigma_2 = 1 + (\theta_{L1}/T_{K2})(T_{L1} - T_{L2})$. In particular, if there is no technical progress in sector 1, the equivalence of the two types of neutralities in sector 2 re-

[26] Substitute (23) into (29) to yield: $\hat{p} = (1 - \theta_{K1}/\theta_{K2})\hat{w} + \theta_{K1}(T_{K2} - T_{K1})$. Thus $\hat{p}\big|_{\hat{w}=0} \gtreqless 0$ according as $T_{K2} \gtreqless T_{K1}$.

[27] If $\pi_2 = 0$, the two neutralities are equivalent in sector 1 only if $\sigma_1 > 1$.

[28] Again, the Hicks rate and bias can be expressed in terms of the u_{ij}^0:

$$\begin{aligned}\pi_j &= \theta_{Kj}\overset{0}{u}_{Kj} + \theta_{Lj}\overset{0}{u}_{Lj} \\ \beta_j &= \overset{0}{u}_{Lj} - \overset{0}{u}_{Kj} + \sigma_j T_{Kj}\end{aligned}$$

quires the elasticity of substitution to be less than unity in that sector.

III. *Some Fundamental Results*

This section is devoted to analyzing the fundamental relations among the three types of measures of bias. These relations help characterize the structure of a two-sector model and include the aggregative (one-product) models as special cases.

Let us define $\rho \equiv T_{L1}/(T_{L1}+T_{K2})$. Clearly $0 \leq \rho \leq 1$. If $\rho=0$, no progress occurs in sector 1; if $\rho=1$, no progress occurs in sector 2; and if $0<\rho<1$, both sectors obtain technical change. To bring all three measures of bias together in a simple form, we eliminate σ_j from equations (34) and (46) to get:

$$(50) \quad \beta_j = \rho\gamma_j + (1-\rho)\alpha_j + \rho(T_{K2} - T_{Kj})$$

Several remarks are suggested by this general expression. First, the Hicksian measure of bias is a weighted average of the other two measures in the consumption-goods sector. This is also true in the capital-goods sector if the Solow rates of progress, T_{Kj}, are equal as between sectors. Second, if there is no progress in the capital-goods sector ($\rho=0$), the character of progress according to Hicks' classification is always equivalent to that of Harrod. In this case, the relation between Hicks' and Harrod's measures is independent of Solow's measure. However, there is no symmetrical result for the mirror-image case. When there is no progress in the consumption-goods sector, Hicks and Solow neutralities can never be equivalent in the other sector. Instead, Hicks neutral must be accompanied by Solow labor-saving and Solow neutral must come with Hicks capital-saving bias. Third, if progress occurs in the capital-goods sector ($\rho>0$: any two neutralities imply the neutrality of the third one in sector 2, and the elasticity of substitution is thus necessarily unity. Fourth, whenever Solow neutrality prevails in sector 2, Hicks' classification is always equivalent to Harrod's in that sector. In the case where Harrod neutrality prevails in sector 2, Hicks' classification (except for the neutrality case)[29] necessarily implies the corresponding classification of Harrod in that sector, but not vice versa. Moreover, if progress is Hicks-neutral in sector 2, Harrod labor-saving (capital-saving) progress necessarily implies Solow capital-saving (labor-saving) improvement, but again not vice versa. Fifth, consider the case where the Solow rates of progress, T_{Kj}, are equal as between sectors.[30] Then the presence of any two types of neutralities implies the third in sector 1. Furthermore, if progress is Harrod (Solow) neutral in sector 1, then Harrod labor-saving, neutral, or capital-saving improvement is equivalent, respectively, to Solow capital-saving, neutral, or labor-saving improvement. If the Solow rates of progress are unequal as between sectors, clearly the differential rate $(T_{K2}-T_{K1})$ is a key factor in linking the three types of measures together. In particular, if sector 1 is more progressive than sector 2 in Solow's sense $(T_{K1}>T_{K2})$, then Harrod (or Solow) labor-saving (or Hicks capital-saving) progress is bound to occur provided that the other two types of progress are both neutral. Lastly, equation (50) embodies the one-product economy as a special case. In that case, it becomes $\beta=\theta_L\alpha+\theta_K\gamma$, or equivalently, $\beta=u_L-q_K$ (here, of course, we delete the subscripts denoting sectors).[31]

[29] Note that the presence of Harrod and Hicks neutralities in sector 2 does not necessarily imply Solow neutrality. This is true only if progress occurs in sector 1 or $\sigma_2=1$. In the latter case, the three neutralities coincide.

[30] This implies that progress occurs in both sectors and that $0<\rho<1$. This case of equal Solow rates of progress as between sectors is equivalent to equal Hicks rates of progress (π_j) as between sectors only if the sectoral factor proportions are equal.

[31] This is the condition obtained by Fei and Ranis (1965, p. 194, equation 4.4) Their B/ϵ, U, and D correspond to β, $-u_L$, and $-q_K$, respectively, in the present paper.

In passing, let us obtain another general expression linking the three own measures of bias together. Eliminating σ_j from (48) and (49) and simplifying yields:

$$(51) \quad \beta_j = \theta_{Kj}\overset{0}{\gamma_j} + \theta_{Lj}\overset{0}{\alpha_j} + \theta_{L1}(T_{L1} - T_{Lj})$$

The Hicksian bias is seen to be a weighted average of the other two own biases in sector 1. This is also true in sector 2 if the Harrodian rates of progress, T_{Lj}, are equal as between sectors. Clearly, the presence of any two neutralities must carry the third in the capital-goods sector while this is the case in the consumption-goods sector only if $T_{L1}=T_{L2}$. Moreover, own neutrality in Harrod's or Solow's sense would imply the equivalence of the other two types of classifications in sector 1 and would also be true for sector 2 only if the Harrodian rates of progress were equal as between sectors.[32] If the Harrodian rate of progress is higher in sector 2 than in sector 1, one may conclude that Harrod's or Solow's own labor-saving (or Hicks capital-saving) progress in sector 2 would occur if the other two types of progress were both neutral.

IV. *Aggregate Neutrality*

In this last section, let us discuss the conditions for obtaining overall neutrality when each sector exhibits a certain kind of neutrality. As William Fellner has suggested, the nature of technical progress in the economy as a whole is not merely a technological matter; it is also influenced by demand. The precise conditions under which aggregation from Hicks and Harrod neutrality in each sector to Hicks and Harrod neutrality for the economy, respectively, have been found by Jones (1965a, 1966). Given the demand assumption that community taste patterns are homothetic, Jones has shown that, aside from the trivial case in which factor proportions are equal as between sectors, aggregation from Hicks (Harrod) neutrality in each sector to Hicks (Harrod) neutrality for the economy follows if either the elasticity of substitution between commodities on the demand side is unity or the Hicksian rates of technical progress, π_j, (the Harrodian rates of progress, T_{Lj},) are equal as between sectors. One might conjecture that analogous conditions would be obtained if we consider the same type of question for Solow neutrality. However, this is only partly correct. To show this, let us first define technological change as neutral for the economy in Solow's sense if per capita income (Y/L) is unaltered at a given real wage rate (measured in consumption goods). It is easy to express the overall labor-income ratio in terms of the sectoral labor-output ratios:

$$(52) \quad L/Y = \theta_1 L_1/pY_1 + \theta_2 L_2/Y_2,$$

where $Y=pY_1+Y_2$, and θ_j represents the share of sector j in the national income (Y). Let $e_L=\hat{Y}-\hat{L}$ and $\lambda_{Lj}=L_j/L$.[33] Differentiating (52) and making use of equations (8), (23), (29), and (45), we obtain:

$$(53) \quad e_L = C_1\hat{w} + C_2 + C_3,$$

where

$$C_1 \equiv \lambda_{L2}\sigma_2 + \lambda_{L1}[(\sigma_1 - 1)\theta_{K1}/\theta_{K2} + 1]$$
$$C_2 \equiv \lambda_{L1}\theta_{K1}\gamma_1 + \lambda_{L2}\theta_{K2}\gamma_2$$
$$C_3 \equiv \lambda_{L1}\theta_{K1}(T_{K2} - T_{K1}) - (\lambda_{L1}\hat{\theta}_1 + \lambda_{L2}\hat{\theta}_2)$$

Given that progress is Solow-neutral in each sector ($\gamma_1=\gamma_2=C_2=0$), aggregate Solow neutrality ($e_L|_{\hat{w}=0}=0$) follows if, and only if, $C_3|_{\hat{w}=0}=0$.

To close the general equilibrium model, the demand side must be introduced. Assume that the ratio of commodities demanded depends only upon relative commodity prices. We have

$$(54) \quad \hat{Y}_2 - \hat{Y}_1 = \sigma_D\hat{p},$$

[32] Hicks neutrality in sector 1 would imply that Harrod's own labor-saving, neutral, or capital-saving improvement is equivalent to Solow's own capital-saving, neutral, or labor-saving improvement.

[33] The θ and λ notation has been used by Amano (1964b) and Jones (1965a, 1966).

where σ_D is the elasticity of substitution of demanders. The introduction of the demand side of the model enables us to determine the behavior of the shares of each sector in the national income:

$$(55)\qquad \begin{aligned}\hat{\theta}_1 &= \theta_2(1-\sigma_D)\hat{p}\\ \hat{\theta}_2 &= -\theta_1(1-\sigma_D)\hat{p}\end{aligned}$$

Substituting (55) into (53) and noting that $\hat{}\,|_{\hat{w}=0}=\theta_{K1}(T_{K2}-T_{K1})$,[34] we find

$$(56)\qquad C_3|_{\hat{w}=0} = \theta_{K1}(T_{K2}-T_{K1})C_4,$$

where

$$C_4 \equiv \lambda_{L1} - (\lambda_{L1}\theta_2 - \lambda_{L2}\theta_1)(1-\sigma_D)$$

Aggregate Solow neutrality obtains[35] if, and only if, $T_{K2}=T_{K1}$ or $C_4=0$. The latter condition can be simplified to:

$$(57)\qquad \sigma_D = 1/(1-\theta_{L1}/\theta_L),$$

where θ_L is the share of labor in the national income. Clearly σ_D is positive if, and only if, $\theta_L>\theta_{L1}$, which is equivalent to the condition that the capital-goods sector be more capital-intensive than the consumption-goods sector.[36] Furthermore, σ_D is positive only if it is greater than unity. We conclude that aggregate Solow neutrality follows if, and only if, either (i) the Solow rate of technical advance, T_{Kj}, is the same in each sector, or (ii) σ_D is greater than unity (equivalently, the consumption-goods sector is less capital-intensive than the capital-goods sector).[37]

As a final remark, it is noteworthy that the completely analgous conditions as between the Harrodian and Hicksian schemes cannot be applied without care to the Solovian one. The only analogous condition is the one which requires the corresponding (Hicksian, Harrodian, and Solovian) rates of progess to be equal as between sectors. This is the condition which ensures both the share of each good in the national income and the ratio of commodity prices to remain unchanged in the process of aggregation in each schema. On the other hand, if the Solow rates of progress differ as between sectors, both the unitary demand elasticity of substitution condition and the equal factor proprotions condition cannot be valid in the Solow scheme although they are both valid in the other two. As is clear from previous discussions, unitary elasticity of factor substitution in each sector and equal Solow rates of progress as between sectors will permit the coexistence of all the three neutralities at each sectoral level. Suppose this to be the case. Then a comparable result at the economy-wide level can be obtained only if factor proportions and all the Hicks, Harrod, and Solow rates of progress are equal as between sectors. Given that factor proportions are not identical in both sectors, we conclude that there can never be a chance of coexistence in neutrality among the three at both disaggregated and economy-wide levels.

REFERENCES

A. Amano, (1964a) "Biased Technical Progress and A Neoclassical Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1964, *78*, 129–38.

———, (1964b) "Determinants of Comparative Costs: A Theoretical Approach," *Ox-*

[34] See fn. 26.

[35] One may still obtain aggregate neutrality with proper offsetting biases at the sectoral levels. For example, γ_1 and γ_2 may not be zero but still C_2 and $C_3|_{\hat{w}=0}$ each or their sum is zero. The major concern here is to consider aggregating neutrality in each sector to the economy as a whole.

[36] The following relation can be derived:

$$1-\theta_{L1}/\theta_L=\theta_2\theta_{K1}(\theta_{L1}/\theta_L)(k_1/k_2-1)$$

[37] It is understood that the possibility that $\sigma_D<0$ is ruled out in stating condition (ii). One may note that if the factor proportion is the same in each sector, the expression $(\lambda_{L1}\theta_2-\lambda_{L2}\theta_1)$ in C_4 of equation (56) would go to zero, and C_4 would not be zero for whatever the values of σ_D. This implies that if $T_{K2}\neq T_{K1}$, the same factor proportion in each sector would not lead to aggregate neutrality.

ford Econ. Pap., Nov. 1964, *16*, 389–400.

———, "Induced Bias in Technological Progress and Economic Growth," *Econ. Stud. Quart.*, Mar. 1967, *17*, 1–17.

K. J. Arrow, H. B. Chenery, B. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

A. Asimakopulos, "The Definition of Neutral Inventions," *Econ. J.*, Dec. 1963, *73*, 675–80.

——— and J. C. Weldon, "The Classification of Technical Progress in Models of Economic Growth," *Economica*, Nov. 1963, *30*, 372–86.

M. J. Beckmann and R. Sato, "Aggregate Production Functions and Types of Technical Progress: A Statistical Analysis," *Amer. Econ. Rev.*, Mar. 1969, *59*, 88–101.

E. Burmeister and R. Dobell, "Disembodied Technical Change with Several Factors," *J. Econ. Theor.*, June 1969, *1*, 1–8.

——— and ——— *Mathematical Theories of Economic Growth*, New York 1970, ch. 5.

P. A. Diamond, "Disembodied Technical Change in a Two-Sector Model," *Rev. Econ. Stud.*, Apr. 1965, *32*, 161–68.

E. M. Drandakis and E. S. Phelps, "A Model of Induced Invention, Growth, and Distribution," *Econ. J.*, Dec. 1966, *76*, 823–40.

J. C. H. Fei and G. Ranis, "Innovation, Capital Accumulation and Economic Development," *Amer. Econ. Rev.*, June 1963, *53*, 283–313.

———, "Innovational Intensity and Factor Bias in the Theory of Growth," *Int. Econ. Rev.*, May 1965, *6*, 182–98.

W. J. Fellner, "Appraisal of the Labor-Saving and Capital-Saving Character of Innovations," F. A. Lutz and D. C. Hague, eds., *The Theory of Capital*, London 1961.

C. E. Ferguson, "Neoclassical Theory of Technical Progress and Relative Factor Shares," *Southern Econ. J.*, Apr. 1968, *34*, 490–504.

F. H. Hahn and R. C. O. Matthews, "The Theory of Economic Growth: A Survey," *Econ. J.*, Dec. 1964, *74*, 779–902.

R. F. Harrod, *Towards a Dynamic Economics*, London 1956.

J. R. Hicks, *The Theory of Wages*, 2nd ed. London 1963.

R. W. Jones, (1965a) "The Structure of Simple General Equilibrium Models," *J. Polit. Econ.*, Dec. 1965, *73*, 557–72.

———, (1965b) "'Neutral' Technological Change and the Isoquant Map," *Amer. Econ. Rev.*, Sep. 1965, *55*, 848–55.

———, "Comments on Technical Progress," *Philippine Econ. J.*, 1966, *5*, 313–32.

C. Kennedy, "Technical Progress and Investment," *Econ. J.*, June 1961, *71*, 292–99.

———, (1962a) "Harrod on 'Neutrality'," *Econ. J.*, Mar. 1962, *72*, 249–50.

———, (1962b) "The Character of Improvements and of Technical Progress," *Econ J.*, Dec. 1962, *72*, 899–911.

J. E. Meade, *A Neo-Classical Theory of Economic Growth*, London 1961.

E. S. Phelps, *Golden Rules of Economic Growth*, New York 1966, Part I Essay 2.

J. Robinson, "The Classification of Inventions," W. Fellner and B. F. Haley, eds., *Readings in the Theory of Income Distribution*, Philadelphia 1949, pp. 175–80.

———, *Essays in the Theory of Economic Growth*, New York 1962.

H. Rose, "The Condition for Factor-Augmenting Technical Change," *Econ. J.*, Dec. 1968, *78*, 966–71.

W. E. G. Salter, *Productivity and Technical Change*, London 1960, ch. 3.

P. A. Samuelson, "Parable and Realism in Capital Theory: The Surrogate Production Function," *Rev. Econ. Stud.*, June 1962, *29*, 193–206.

R. Sato and M. J. Beckmann, "Neutral Inventions and Production Functions," *Rev. Econ. Stud.*, Jan. 1968, *35*, 57–67.

R. M. Solow, "Technical Progress, Capital Formation and Economic Growth," *Amer. Econ. Rev.*, May 1962, *52*, 76–86.

———, *Capital Theory and the Rate of Return*, Amsterdam 1963.

A. Takayama, "On a Two-Sector Model of Economic Growth with Technological Progress," *Rev. Econ. Stud.*, July 1965, *32*, 251–62.

H. Uzawa, "Neutral Inventions and the Stability of Growth Equilibrium, "*Rev. Econ. Stud.*, Feb. 1961, *28*, 117–24.

Measuring Economic Stabilization: 1955–65

By WAYNE W. SNYDER*

A recent study by Bent Hansen (1969) gives the institutional background to budgetary action and an analysis of the nature and effect of fiscal policy for each of seven *OECD* countries: Belgium, France, Germany, Italy, Sweden, the United Kingdom, and the United States. The Hansen study uses three measures to characterize the impact of budget changes: the average annual effect on domestic demand, the effect on the secular rate of *GNP* growth, and the short-term stabilization around the trend of *actual* GNP. The study did not measure their effects in relation to *potential* (i.e., full employment) *GNP*. Partly, this was because there were no officially recognized estimates of the level of domestic demand appropriate to maintain full employment growth for each country. For three of these countries—Sweden, the United Kingdom, and the United States—estimates of potential output now exist, and my purpose is to supplement the Hansen study by comparing the total impact of budget changes with these *optimum* levels of *GNP*.

I. *Measuring the Impact of Budget Changes*[1]

This study measures the combined budgetary impact on domestic demand of both automatic and discretionary effects of budget changes.[2] This includes the direct impact brought about by the initial budget change as well as the subsequent indirect or "multiplier" effect.[3] The definitions and methods employed to estimate the total effect are the ones developed by Hansen and are based on earlier contributions of E. Cary Brown, Bent Hansen (1959), Assar Lindbeck (1956), and Richard Musgrave. Although the Hansen model is small compared with the large econometric models which have been developed for some countries, it is adequate to measure the relative importance of budget changes between countries. Due to the lack of quarterly data for all seven countries, the model uses only year-to-year changes, and there are no lags.[4] The model assumes that private investment and exports are exogenously determined. Imports are endogenous and for some countries, e.g., Sweden, represent the

* Center for Research on Economic Development, University of Michigan. This article was suggested by J. C. R. Dow, who first pushed for an *OECD* evaluation of fiscal policies. More fundamentally, however, I am indebited to Bent Hansen for the two very rewarding years that I assisted him on the study. The final version benefitted from their comments on an earlier draft, and from many useful suggestions of my colleagues Harold Shapiro and Lester Taylor, a referee, and the editor. Janet Eckstein helped by proposing several editorial changes.

[1] For a complete description of the methods used to measure budgetary effects, see Hansen 1969, ch. 1.

[2] In a growing economy where the "fiscal drag" from automatic tax increases can be important, albeit compensated to varying degrees by built-in expenditure programs with expansionary effects, narrowly defined discretionary effects may be the appropriate measure if short-term stabilization is the primary interest. The Hansen study does, in fact, provide separate estimates for both automatic and discretionary effects, but my concern here is with the budget's *total* impact on achieving balanced growth.

[3] An "accelerator" effect should be included too, but (as will be explained later) the actual model assumed that all changes in private investment were exogenously determined; hence the measurement of the budgetary impact is limited in this respect as well as by the others described further on.

[4] A review of several big models suggests that three-fourths or more of the budget effect generally occurs during the first year, hence the absence of explicit lags is not critical. See Hansen (1969, pp. 20–22) for a discussion of this subject.

principal leakage of potential budget effects.

The Hansen methodology for estimating automatic and discretionary budget effects separately, does include automatic tax response rates but these are not necessary to calculate the combined budgetary impact, or the "total effect" of budget changes. Hansen's formula for this is based on a truncated version of his model and allows for all changes in revenue which are not credit transactions and for all purchases of goods and services, except direct government imports which are excluded because they do not affect domestic demand. Thus, the only formula used in this study is the following:

$$\text{Total Effect} = \frac{1}{1-\alpha(1-\mu)}[(dg+ds) + \alpha(1-\mu)(gdp_g + sdp_s) - (1-\mu)dT_i - \alpha(1-\mu)dT_d]$$

where α is the marginal propensity to consume; μ is the marginal propensity to import with respect to GNP; dg and ds are annual changes in the volume of goods and services, respectively, purchased by government; gdp_g and sdp_s are changes in the value of goods and services due to price or wage changes; dT_i and dT_d are annual changes in indirect and direct personal taxes. The multipliers for the various types of budget changes differ of course between countries because of differences in the leakage coefficients—the marginal propensity to consume and, especially, the marginal propensity to import. Leakage coefficients and multipliers are given in Table 1.

If the consumption coefficients seem small, this is because α is the ratio of changes in personal consumption to changes in *total* private income minus only direct household taxes, a definition required because the model does not include an explicit corporate sector.[5] The expenditure multipliers may seem large, but it should be noted that they refer to budget effects excluding tax leakages which are accounted for by explicitly including dT_i and dT_d (scaled by their appropriate multipliers, too). The corresponding multipliers which include tax leakage coefficients (instead of dT_i and dT_d) are substantially lower; for example, the multiplier for changes in the volume of government purchases of domestically produced goods and services in the United States is 3.48 *without* tax leakages but only 2.12 *after* allowing for normal tax increases. While the multipliers cannot be accepted as being exact or applicable for every budget change, they are, nevertheless, sufficiently representative to indicate relative orders of magnitude and the range of differences between countries.

The choice of which governmental sector's budget changes to include (i.e., central, state-local, social security, public enterprise investment) was based on the budget policies that seemed to be substantially influenced by the central government. On this basis, budget changes

TABLE 1—LEAKAGE COEFFICIENTS AND MULTIPLIERS

	Leakage Coefficients		Multipliers for Total Effects		
	α	μ	dg $+ds$	gdp $+sdp_s$ and dT_d	dT_i
Sweden	0.80	0.40	1.92	0.92	1.15
United Kingdom	0.68	0.28	1.96	0.96	1.41
United States	0.75	0.05	3.48	2.48	3.30

Source: Hansen (1969, pp. 46–47).

[5] All models have their deficiencies, and Hansen's is no exception. In this respect, perhaps the most conspicuous features are that private investment is treated as an exogenous variable and that the corporate sector is not made explicit. The first can be explained by the inadequate knowledge about investment functions, especially for European countries; the second was necessary to maintain comparability among the original seven countries, because the national accounts do not all give corporate profits and taxes separate from personal income and taxes, notably Sweden for which business savings contain an important error item.

include the general government plus the investments of public enterprises for Sweden and the United Kingdom, but only the federal government (including the federally financed portion of the social security sector) for the United States.

II. *Evaluating Economic Stability and Growth*

The key quantities considered in this study are actual and potential *GNP*, and the total effect of budget changes. Figure 1 shows for each country how actual *GNP* (in constant 1958 prices) developed between 1955 and 1965 in relation to potential output. Official estimates of potential *GNP* are available for two countries: W. A. H. Godley and J. R. Shepherd's estimates are used for the United Kingdom, and those made by the Council of Economic Advisors (Jan. 1970) are used for the United States. Erick Lundberg's estimates are used for Sweden. Estimates of potential output are, of course, somewhat problematic because they are not independent of economic policy; government policy affects the distribution of output between consumption and investment, and this clearly influences the rate of growth of the labor force and productivity. But in spite of these limitations, they provide a useful basis for evaluating budgetary performance if they are discussed in relation to other policy objectives and instruments.

The total effect of budget changes should not be directly compared with actual and potential *GNP* because actual *GNP* is itself influenced by budget changes. We can, however, construct a hypothetical series of *GNP* by subtracting from actual *GNP* the total effect for each year. This derived series is called the "pure cycle," because it attempts to estimate what *GNP* would have been each year without the budgetary impact. The pure cycle still incorporates the effects of other govern-

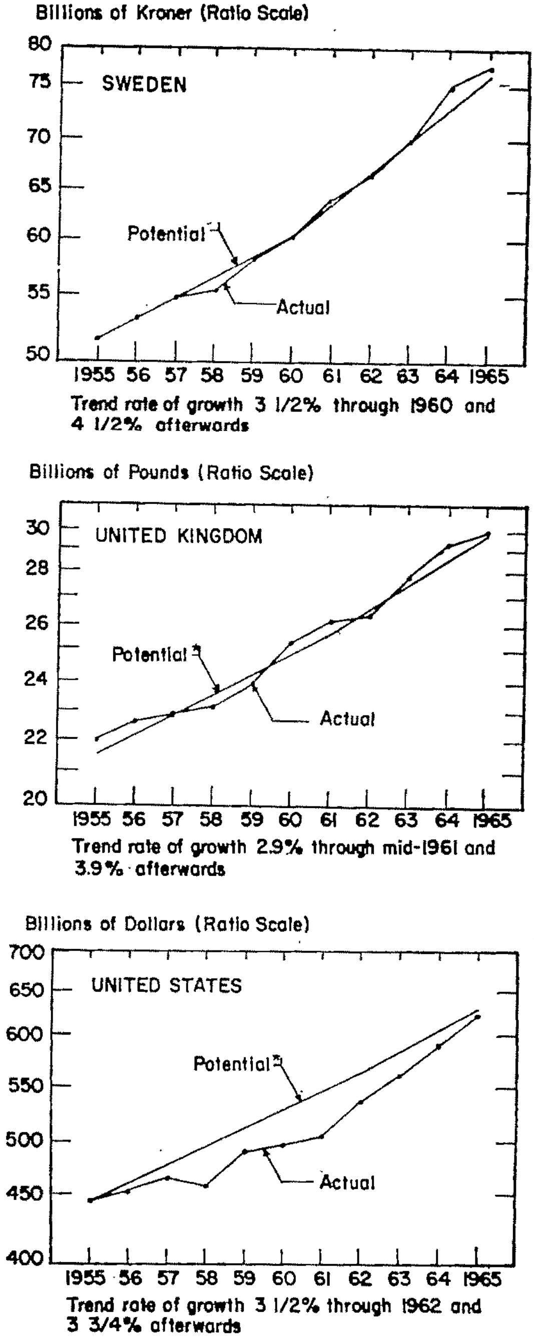


FIGURE 1.—GROSS NATIONAL PRODUCT: ACTUAL AND POTENTIAL (Constant 1958 Prices)

ment policies (e.g., monetary and direct controls) and autonomous forces (e.g., private investment and exports). Hence, the pure cycle is not so pure, but nevertheless it is a useful analytical construct.

In Figures 2 (Sweden), 3 (United Kingdom), and 4 (United States), actual *GNP* and the pure cycle are expressed as percentage deviations from potential output, shown for convenience as a horizontal line. The vertical difference between actual *GNP* and the pure cycle for each year is equivalent to the estimated total effect of budget changes. The arrows indicate the direction of the total effect and point from the pure cycle to the actual *GNP*.

The *GNP* development given in Figure 2 for Sweden is different than for the two other countries because in seven out of the eleven years, actual *GNP* was virtually identical with potential output, and this required dampening as well as expansionary budget policies, with annual total effects which amounted to 2 and 3 percent of potential *GNP*. Perhaps this is not surprising, since for several decades fiscal policies have been accepted and employed to manage the Swedish economy, for which Lindbeck (1968) points out that the first "... deliberate counter-cyclical fiscal policies [were] introduced as early as 1933" (p. 33).

The extraordinary flexibility and power of fiscal policy when properly employed is illustrated by the Swedish experience of 1959 and 1960. Following the worldwide recession in 1958 which budget policies helped partially eliminate, the total effect of budget changes gave an expansionary push of 3 percent which helped carry demand up to potential output. As the recovery progressed, the budget policies were reversed. For 1960, the budget switched to a dampening of 3 percent—just the amount needed to keep the economy at virtually its full potential for both 1959 and 1960. The only mistakes occurred

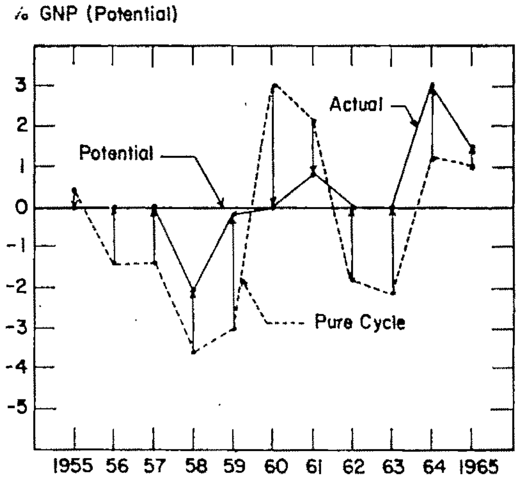


FIGURE 2.—SWEDEN

in 1964 and 1965, when expansionary policies fueled an already overheated economy, which was brought back into balance only after 1965.

As for the other main economic objectives, Sweden did not have any serious problems concerning the balance of payments equilibrium, but this was not true of relative price stability. Current account difficulties occasionally required some minor adjustments of domestic policies, but over the period as a whole Sweden ran a sufficiently large surplus on its basic account to permit doubling its international reserves between 1955 and 1965. Prices increased on average more in Sweden than in either the United Kingdom or the United States, but not as much as for some other European countries. Lundberg suggests that "... a serious criticism of the type of stabilization policy pursued in Sweden is that during most of the time it involved a combination of inflationary pressure and extensive controls of the credit and capital markets" (p. 199) while at the same time "... there has been no effort by the Swedish government to carry out an 'income policy' or interfere with the bargaining process on the labor market" (p. 248). If these considerations should

have received higher priority, then clearly greater slack in the labor market should have prevailed, which means that potential output would have been lower. But even if a somewhat higher level of unemployment was used to calculate potential *GNP*, the stabilization achieved would still remain impressive. Furthermore, Sweden used a very sophisticated and effective method to provide investment funds to firms in a countercyclical manner; if quantitative estimates of these policies were made, the picture of Sweden's already impressive budgetary performance would be further improved.[6] In sum, Sweden during all but the last two years of the period between 1955 and 1965 presents a remarkable picture of the effectiveness of using budgetary policies to achieve economic stability and growth.

The United Kingdom's budgetary performance is striking by comparison with Sweden's. The requirements for domestic demand management were similar in both countries, but U.K. policymakers did very little to alter the course of events. Except for the years 1955 and 1965, the total effect of budget changes never greatly exceeded 1 percent of potential output; thus, in spite of its "stop/go" policies, budget changes had little influence on the underlying pattern of cyclical fluctuations. It is particularly important, however, to recall the lack of purity in the pure cycle, which includes the effects of monetary policies; for example, variations in down-payments and the length of maturity requirements for consumer credit on durable goods were especially important at various times.

Unlike the other six countries in the Hansen study, the United Kingdom accentuated rather than reduced the gap between potential and actual *GNP* during the 1958 recession. Prices had been rising sharply for several quarters, and policies were designed partly to counteract this and partly to discourage further speculation against sterling, such as the short-lived crisis of 1957. The result was, a reduction in domestic demand at a time when unemployment was already rising to what in England was considered an intolerable level.[7]

This experience is typical of the problems that plagued the United Kingdom throughout this period. The definitive analysis of the British postwar experience, the low rate of growth (lowest among the seven countries in the Hansen study), how much higher a rate might have been attained if policies had been different, the many balance of payments crises, and the problems with rising prices, is still to be written. Until this is done, J. C. R. Dow's study will remain the most complete analysis, in which he asserts that ". . . the major fluctuations in the rate of growth of demand in the years after 1952 were thus chiefly due to government policy" (p. 384). And he also concludes that: "If the pressure of demand had been somewhat lower and the margin of capacity somewhat larger . . . the pressure of demand could have had a marked effect on the rise of prices (p. 361) . . . and would probably have reduced fluctuation in the balance of payments . . . " (p. 392).

The minimum that needs to be said regarding these propositions is that *if* the potential output used is considered to have been too optimistic as regards price and balance-of-payments developments, given

[6] See Lundberg (pp. 225–32) and Gunnar Eliasson for descriptions and evaluations of this scheme.

[7] Samuel Brittain has described the relation between employment and the balance of payment restraint prior to 1965 in the following way:

> If unemployment (after allowing for the purely seasonal element) is down to 1.5% the balance of payments will usually be given priority; and the Treasury would not be deterred from depressing home demand by the thought of unemployment rising to say 1.8%. But if the number out of work is much higher than this at the time of the Budget, risks may sometimes be taken with the balance of payments for the sake of domestic expansion. [p. 105]

actual budgetary policies, then lowering it would make the policies even less appropriate. And this conclusion is not affected by the exclusion of variations in corporate taxes and the various investment allowance schemes which were used throughout the period. Corporate tax payments lag as much as two years behind the time that profits are being earned, and hence tend to produce pro rather than counter cyclical effects, and when discretionary changes in corporate tax rates and changes in the system of investment allowances are combined, they produce an even greater destabilizing impact on domestic demand.

If differences between Sweden and the United Kingdom as illustrated in Figures 2 and 3 are striking, the United States presents yet another situation. The problem of economic management was not the fine tuning of demand at full employment which characterized both Sweden and the United Kingdom, but rather dealing with the increasing tendency of demand to diverge from potential output after 1955. The annual budget impacts unquestionably helped to dampen fluctuations of actual *GNP*, but they did relatively little to assist the economy to regain full employment. Figure 4 makes it clear that the counter cyclical effects of budget changes

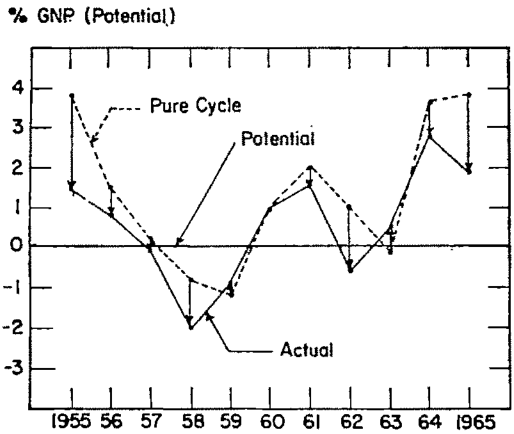


FIGURE 3.—UNITED KINGDOM

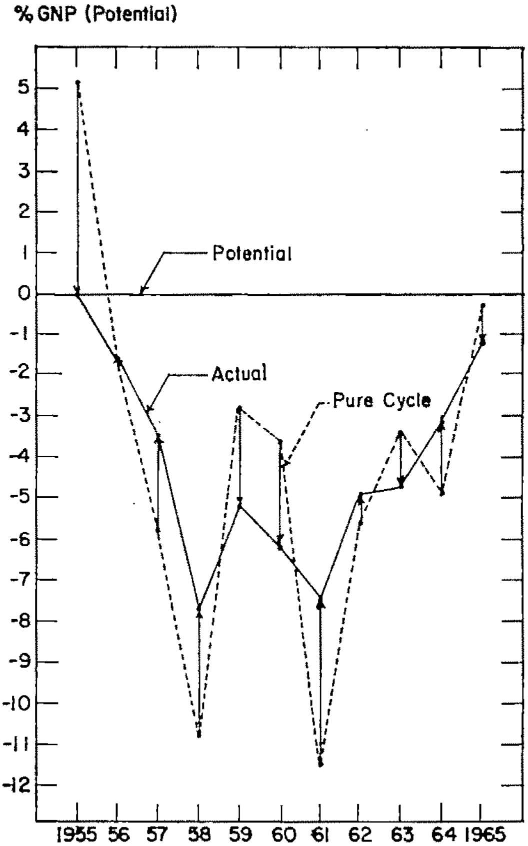


FIGURE 4.—UNITED STATES

were particularly—and substantially—inappropriate during 1959 and 1960; the same is true for 1963 and 1965, albeit to a lesser degree. Only in 1955 were the effects entirely adequate. The destabilizing effects in years of post-recession recovery are mainly due to the strong automatic stabilizers, as noted by Wilfred Lewis; although they greatly help to limit the severity of recessions, they reverse direction quickly after the trough is passed and hinder recovery to full employment.

Balance-of-payments equilibrium and relative price stability are other objectives which need to be considered in discussing economic stability and growth, but neither is as important for the United States as for the United Kingdom. Toward the end of

the 1950's the *U.S.* policymakers became increasingly concerned about the continual balance of payments deficits, but no specific budget policies for these purposes were adopted at any time during the period from 1955 through 1965, although some monetary policies were directed to the objective of improving the balance of payments. In presenting the 1963 tax reduction proposals, it was even suggested that the subsequent expansion toward full employment might eventually help the balance of payments by encouraging American companies to invest less overseas and more at home.[8]

Price considerations did, however, play a more important role but to a large extent they are inseparable from the assumption that a 4 percent level of unemployment is compatible with potential output and with little or no inflationary pressure. While the Eisenhower administration did not specify what trade off it thought existed between price stability and unemployment, clearly it recognized a linkage and was more concerned about the rising consumer prices from 1955 until 1959, which were fairly large by American standards (on average about 2 percent annually), than it was about reducing the level of unemployment to 4 percent. The Kennedy administration did make explicit its goals and did believe that it was possible to reduce unemployment to at least 4 percent while maintaining relative price stability. Perhaps it would be more reasonable to propose a potential *GNP* corresponding closer to 5 percent unemployment for the years prior to 1961. If this were done the amount of divergence between the pure cycle and potential *GNP* would be smaller and consequently the amount of stabilization achieved would be somewhat greater. The importance of this should not, however, be overemphasized; even with such an adjustment, the *U.S.* economy functioned well below its potential throughout most of the period.

It is not a major purpose of this study to explain why policies were inadequate but it can be noted that throughout the Eisenhower administration there was a general feeling that the economy would reach an appropriately high level of economic activity without the discretionary use of government policies. And as late as January 1962 the new Kennedy administration believed that the economy could achieve a 4 percent level of unemployment by mid-1963 without any major discretionary policies beyond the investment credit scheme and accelerated depreciation allowances which it proposed to increase investment and productivity.[9] Eventually, of course, the substantial tax cuts of 1964 and 1965 were required in order to boost the economy toward potential *GNP*, and not until 1966 did expenditures related to the Vietnam war finally raise the economy to its full potential, after a decade of inadequate demand.

III. *Comparing Relative Stabilization*

Although important issues of methodology and interpretation which cannot be neglected are raised in this section, a tentative and quantitative evaluation of the relative degree of stabilization achieved by the three countries seems appropriate. First some definitions, and then the conceptual problems are discussed.

The absolute difference between the pure cycle and potential *GNP* is defined as *potential* stabilization. Budget effects are counted as stabilizing if they diminish the difference and destabilizing if they increase it. The total effects whose arrows in Figures 2, 3, and 4 point towards potential *GNP* are counted as helping to achieve economic stability, and in those years when they point away the budgetary impact is counted as being destabilizing.

[8] See Council of Economic Advisors (Jan. 1963, p. 103).

[9] See Council of Economic Advisors (Jan. 1962, p. 66 and p. 132).

Budget policies are nowhere uniquely designed for the management of domestic demand; other objectives sometimes override stabilization considerations, nor are budget policies the only means by which objectives can be attained.[10] Consider for example an extreme situation where the budgetary impact is always in the wrong direction, i.e., causing demand to diverge from potential *GNP*. If, however, the budget effects are precisely forecast and if other government policies are used to eliminate exactly whatever difference would exist, *ceteris paribus*, between the pure cycle and potential output, then the economy would always be at its full potential but all the stabilizing effects would be attributed to the budgetary impact rather then the other policies which were the real reason for maintaining full employment.[11] Such possibilities cannot be denied but their importance must be viewed in light of individual experience. In the United States during ten of the eleven years from 1955 through 1965, actual *GNP* remained below—often substantially below—potential output, the cumulated shortfall amounting to nearly 50 percent of a typical year's potential *GNP* (measured in 1958 prices). On the other hand, throughout this period domestic demand in both Sweden and the United Kindgom fluctuated within a narrow band between nearly full and (more often) overfull employment. In these cases the possibility does exist that some stabilizing effects which are attributed to the budgetary impact properly belong to the counter-balancing effects of other government policies. While this study makes no attempt to evaluate the effects of other government policies, it is worth mentioning that Sweden and the United Kingdom used monetary policy mainly for balance-of-payments purposes which had, if anything, adverse rather than stabilizing effects on the appropriate management of domestic demand.

This does not exhaust all the conceptual problems; at least one more should be discussed. In cases where the budgetary impact is in the right direction but exceeds the amount necessary to reach potential output, how should the total effect be calculated? It seemed sensible to divide the impact into two components: the part that was stabilizing, and the other which overshot and had destabilizing effects. In fact, such situations were rare, never occurring in either Sweden or the United States; the only clear case occurred in the United Kindgom in 1962 when the total effect was about twice as strong as necessary to dampen the overheated economy and created some undesirable unemployment.

With these definitions and caveats in mind, we can make the following evaluation of the budgetary impact and its contribution to achieving economic stability in the three countries. Table 2 summarizes the pertinent relationships on a cumulated basis for the eleven-year period.

First, the cumulated amount of potential stabilization (Item 1) for Sweden and the United Kingdom amounted to about 20 percent (of a typical year potential output), while for the United States it amounted to 55 percent. Second, one-third of Sweden's desirable stabilization and most of the United Kingdom's would have required a dampening budget impact whereas for the United States—except for the year 1955—stabilization required expansionary policies.

There were notable differences among the three countries concerning the stabilizing impact of budget changes (Item 2). In Sweden the cumulated impact had a sizeable stabilizing effect, and it occurred when the pure cycle was above potential *GNP* as well as when it was below it.

In the United Kingdom the stabilizing

[10] E. S. Kirshen's study lists eight major conjunctural and structural objectives and four minor targets; and the same study enumerates no fewer than sixty-five instruments available to achieve them.

[11] The referee pointed out the ambiguity caused by this conceptual problem.

TABLE 2—TOTAL EFFECTS AND ECONOMIC STABILIZATION (expressed as a percentage and cumulated for 1955–65)[a]

	Sweden	United Kingdom	United States
1. Total (absolute) divergence between pure cycle and potential *GNP*	21.0	18.6	55.4
a) Above potential	7.7	16.6	5.1
b) Below potential	13.3	2.0	50.3
2. Sum of stabilizing effects	15.7	7.3	16.9
a) Above potential	4.7	7.1	5.1
b) Below potential	11.0	0.2	11.8
3. Sum of destabilizing effects	2.3	2.3	7.3
a) Above potential	2.3	0.5	0.0
b) Below potential	0.0	1.8	7.3
4. Net stabilizing effects (2 minus 3)	13.4	5.0	9.6
a) Above potential	2.4	6.6	5.1
b) Below potential	11.0	−1.6[b]	4.5
5. Total divergence between actual and potential *GNP* (1 minus 4)	7.6	13.6	45.8
a) Above potential	5.3	10.0	0.0
b) Below potential	2.3	3.6	45.8
6. Net stabilization achieved (4÷1)	63.8	26.9	17.3

[a] Budgetary effects were measured for general government plus the investments of public enterprises for Sweden and the United Kingdom, but only for the federal government of the United States; the choice was based on what budget policies were substantially influenced by the central government.

[b] Minus sign indicates destabilizing effects.

effects were not very large and were concentrated almost entirely in dampening the potentially overheated economy. In the United States the expansionary effects were about as large as in Sweden, but since the cumulated shortfall of the pure cycle below potential *GNP* was so much greater (50 percent for the United States as compared with only 13 percent for Sweden), the relative amount of stabilization achieved was substantially less.

No country escaped having destabilizing effects in some years, but they were relatively small for both Sweden and the United Kindgom and somewhat larger for the United States (Item 3). All of Sweden's destabilizing effects contributed to pushing actual GNP above its potential. In the United States the situation was exactly reversed, as the destabilizing impact of budget changes contributed to increasing the gap between potential and actual *GNP*. The "stop/go" policies of the United Kingdom are reflected in the destabilizing effects, which sometimes caused overheating and sometimes contributed to depressing *GNP* below its potential level.

When the stabilizing and destabilizing effects are combined to obtain the net effects (Item 4), and when these are compared with the potential stabilization which could have been achieved (Item 1), there are striking differences between the three countries. For Sweden, the potential stabilization which could have been achieved was not very large to begin with, but 64 percent of it was accomplished; a remarkable achievement by itself but particularly important when compared with the mediocre performance of the other two countries. The potential stabilization which could have been achieved in the United Kingdom was smallest among the three countries, and in fact the effects of budget changes were no larger than those of Sweden; the net result was that 27 percent of the potential stabilization was achieved. The United States had two and one-half to three times as much potential stabilization to be achieved, but the net stabilizing effects of its budget changes were relatively small—only 17 percent.

The cumulated divergence between actual and potential *GNP* for the entire eleven-year period (Item 5) was smallest for Sweden (only 7.6 percent), somewhat larger for the United Kingdom (13.6 percent), and much larger for the United States (45.8 percent). For both Sweden and the United Kingdom, most of the divergence arose from an overheated econ-

omy where actual *GNP* exceeded potential output, although for each country some short-falls did occur. For the United States, however, the entire amount represented the short-fall of actual below potential *GNP*.

IV. *Conclusion*

An important finding of the Hansen 1969 study was that if short-run stabilization is measured with respect to the *actual* trend rate of *GNP* growth, then the degree of stabilization achieved in the United States was substantially larger than for any other country during the eleven years from 1955 through 1965 and was modestly important for Sweden, while in the United Kingdom the impact of budget changes was actually destabilizing and created greater fluctuations in the rate of *GNP* growth than would have occurred if the budget had been neutral from year to year.

If *potential* instead of actual *GNP* is used, the results are strikingly different. The total impact of budget changes in Sweden eliminated nearly two-thirds of the gap between the pure cycle and potential *GNP* and helped create a level of demand that was virtually identical with potential output during a majority of the years. The United Kingdom's performance also is improved, from a generally destabilizing pattern to one where about one-quarter of the potential stabilization was achieved. The change in the United States' performance is less surprising because it is already widely recognized that while the budgetary impact—mainly the automatic built-in stabilizers—helped dampen short-run fluctuations, during the period from 1955 through 1965 the economy slipped below 95 percent of its full employment potential during half of the years. Consequently, the amount of stabilization achieved by this more relevant criteria was only 17 percent.

The reader must, however, interpret these conclusions with caution, keeping in mind the other main economic objectives besides achieving a high level of employment and growth, principally relative price stability and balance of payments equilibrium.

REFERENCES

S. Brittain, *The Treasury Under the Tories 1951–64*, Middlesex 1964.

E. C. Brown, "Fiscal Policy in the Thirties: A Reappraisal," *Amer. Econ. Rev.*, Dec. 1956, *46*, 857–79.

J. C. R. Dow, *The Management of the British Economy, 1945–60*, Cambridge 1964.

G. Eliasson, *Investment Funds in Operation*, Stockholm 1965.

W. A. H. Godley and J. R. Shepherd, "Long-term Growth and Short-term Policy: The Productive Potential of the British Economy and Fluctuations in the Pressure of Demand for Labour, 1951–62," *Nat. Inst. Econ. Rev.*, Aug. 1964, *29*, 26–33.

B. Hansen, "Statsbudgetens verkningar," *Ekonomisk Tidskrift*, 1959:3.

——— assisted by W. Snyder, *Fiscal Policy in Seven Countries, 1955–65*, Paris 1969.

E. S. Kirschen et al., *Economic Policy in Our Time*, Amsterdam 1964.

W. Lewis, *Federal Fiscal Policy in the Postwar Recessions*, Washington 1962.

A. Lindbeck, "Statsbudgetens vernigar pa konjunkturuveckliagen," *SOU*, 1956:48.

———, "Theories and Problems in Swedish Economic Policy in the Post-War Period," *Amer. Econ. Rev., Supp., Part 2*, June 1968, *58*, 1–87.

E. Lundberg, *Instability and Economic Growth*, New Haven 1968.

R. A. Musgrave, "On Measuring Fiscal Performance," *Rev. Econ. Statist.*, May 1964, *66*, 213–20.

J. R. Shepherd, "Productive Potential and the Demand for Labour," *Econ. Trends*, Aug. 1968, ch. 25–27.

Council of Economic Advisers, *Economic Report of the President*, Jan. 1962, Washington.

———, Jan. 1963, Washington.

———, Jan. 1970, Washington.

COMMUNICATIONS

$45 Billion of U.S. Private Investment Has Been Mislaid: Comment

By George Jaszi*

Robert Gordon should be commended on a painstaking statistical effort, notwithstanding some serious flaws in his estimates. According to calculations made at the Office of Business Economics (*OBE*), the value of the investment which he calls "mislaid" is $76 billion rather than $45 billion. The *OBE* estimates of the investment types he discusses were published in a study by Robert Wasson, John Musgrave, and Claudia Harkins. However, I shall not deal here with statistical procedures. Rather, I shall examine the theoretical aspects of Gordon's article.

As a preliminary, I should like to note that Gordon conveys the impression that he has discovered something of which no one before him was aware: that government owned, privately operated capital is omitted from the *OBE* estimates of private investment and private capital stocks. In fact, the omission of government owned, privately operated capital was discussed explicitly in 1956 in a basic *OBE* study devoted to manufacturing investment by Donald Wooden and Robert Wasson. *OBE* brought this study to Gordon's attention long before the publication of his article. All subsequent *OBE* publications on capital stocks have also made clear that they cover only privately owned stocks and exclude government owned capital. Moreover, Victor Perlo developed earlier many of the points featured by Gordon.

If Gordon was nevertheless bent on creating the impression that he has made a discovery, he could with equal justification have staked out a broader claim. He could have discovered that government owned, government operated capital also is omitted from the *OBE* national accounting system, and he could have given his article an even more catchy title by calling it "$768 Billion of U.S. Investment Has Been Mislaid," instead of settling for a measly $45 billion.

But let us discuss Gordon's views on the definitional and conceptual problems underlying the measurement of capital. He writes as though his theoretical framework were obviously and unequivocally correct. In fact, however, he is dealing with difficult matters on which no agreement exists. To discuss these matters, it is necessary to distinguish between two components of capital which Gordon merges in his discussion. One is government surplus capital bought by the private sector; such capital belongs in the private capital stock and the question is how to value it. The other is government owned capital operated by the private sector; the question here is whether such capital belongs in the private capital stock at all.

I shall argue that: 1) within the framework of the national accounting system used by *OBE*, government surplus capital is now properly valued in the *OBE* gross private investment series; 2) Gordon's valuation of government surplus capital as a component of the private gross capital stock is probably too high; 3) inclusion of government owned, privately operated capital in the private capital stock is inappropriate for production function analysis if *OBE* estimates of national product are used to measure output; and 4) that the "where-used" criterion proposed by Gordon for the classification of

* Director of the Office of Business Economics of the U.S. Department of Commerce. He acknowledges the help received from Edward F. Denison and from his associates at the Office of Business Economics. However, they do not agree with all the points made in this paper.

capital stocks cannot be substituted for the "where-owned" criterion now employed.

I believe that Gordon is closest to firm ground with respect to the valuation of government surplus capital bought by the private sector, which accounts for about $7 billion of his $45 billion total. In the *OBE* series, this capital is valued at the price at which it was purchased by the private sector. This can be taken to represent its net (depreciated) value; Gordon argues that it should have been valued at its gross (undepreciated) value. There is a good deal of merit to this argument, simply because the acknowledged aim is to measure gross rather than net capital. Yet I cannot fully agree with Gordon even in this case.

I. *Gross Investment in Government Surplus Capital*

The first point on which I disagree concerns the valuation as part of gross investment of government surplus capital bought by the private sector. This is not a point of great substance. I mention it only to dispel the impression, which may have been created by Gordon's use of the word "mislaid," that there has been an oversight on the part of *OBE*. This is not so; the valuation adopted was intentional. In formulating his position, Gordon forgets that the national accounting system of *OBE*, as now constituted, recognizes tangible capital formation only for business, but not for consumers or for government; and that the definitions that must be adopted, given this constraint, are different from those that would be appropriate if, as he seems to assume, the accounting sytem recognized tangible capital formation for consumers and government as well.

Specifically, it can be shown that, given the truncated form of the *OBE* system, transfers of used capital assets among the sectors must be registered at their net, depreciated values in order to produce consistent entries in the interrelated set of production, appropriation, and saving-investment accounts that make up the system. A gross, undepreciated valuation of these items would result in inconsistent series.

Suppose, for instance, that in a given year government sells surplus capital having a gross, undepreciated value of $10 to business at a net, depreciated value of $4. To simplify matters, assume that the transaction is carried out in cash. There are no other transactions. Clearly, the gross national product must be zero. This is accomplished by *OBE* by registering gross investment at $4 and government purchases at minus $4. Matching this investment of $4, there is government saving of $4, which would be derived as government revenue (zero) less government purchases (minus $4), and would also be equal to the change in the government cash balance. Correspondingly, business saving would be zero (zero income minus zero expense), or investment of $4 offset by a $4 reduction in cash. The alternative, which Gordon would prefer, of registering investment as $10 and government purchases as minus $10, is not open because it would result in an inconsistent accounting for saving. Government saving of minus $10, derived as government revenue (zero) minus government purchases (minus 10), could not be reconciled with the change in cash, because the necessary government depreciation entries are not recognized in the *OBE* system of accounts. And, for similar reasons, there would be an inconsistency in the accounting for business saving. This is the basic reason for using in the gross investment component of *GNP* the valuation of government surplus property sold to business which is criticized by Gordon.

II. *Gross Stocks of Government Surplus Capital*

I have elaborated this point, because Gordon couches his entire criticism in terms of the *OBE* gross investment series. However, it is, as I have noted, not a major point of substance, because his point can be restated in terms of the measurement of gross capital stocks. In terms of our example, he would value gross private capital stock at $10 instead of an *OBE* valuation of $4. Applied to capital stock measurement, his proposal has a good deal of plausibility. After all, it is our acknowledged aim to measure gross rather than net capital stocks. Yet I hesitate

to accept his proposal as the best solution, because I feel that it may result in estimates of the private capital stock that are too high from the standpoint of realistic economic analysis. The series best suited for that purpose lies, in my opinion, between the *OBE* series Gordon criticizes and the one he proposes to substitute.

Suppose that government surplus capital sold to the private sector were less productive in producing output for private use than it had been in producing output for government. Surely, it would be extreme to enter it, as Gordon proposes, at the value paid for it by the government; the contribution of capital to production would be overstated if a series on capital stocks based on such a valuation were used in production function analysis. Instead, the ideal procedure might be to value the surplus capital at the price which private buyers would have been willing to pay for it, had it been new; this is the same as the price which they would have been willing to pay for newly produced capital goods of equal productivity, designed specifically for the civilian use for which the government surplus capital was intended.

While Gordon's valuation may be too high as compared with such a yardstick, *OBE*'s valuation appears to be too low. However, the latter impression is mitigated by the consideration that the value at which the surplus capital should be entered is net of the additional investment that buyers of government surplus capital must make to adapt it to private use. On the basis of the information at my command, I have no firm opinion where, in the range between Gordon's estimates and those of *OBE*, the correct estimates of gross private capital stocks lie. It would seem to me that there is a strong presumption that government surplus capital was less well adapted to production for private use than for government use. This was the part of investment required by the war that business was unwilling to finance privately in spite of generous special amortization and other incentives; in other words, this was the investment judged least likely to succeed in a civilian reincarnation. I know that this consideration does not lead very far, and I do not doubt that opposite considerations could be adduced. But it is not the purpose of these comments to resolve uncertainties. Just the opposite, it is their purpose to show that great uncertainties exist: Things are not as simple as Gordon implies.

One further point should be touched upon before we proceed to the next topic, Gordon's treatment of government owned, privately operated capital. It may be legitimately asked why, if valuation procedures involving a markdown from the initial value of capital goods are justified for government surplus capital sold to the private sector, similar procedures are not entertained for elements of the private capital stock that do not change hands. The answer is twofold. In the first place, we are not aware of shifts in the end use of production likely to affect the productivity of capital that are similar in magnitude to the one associated with government surplus capital after World War II. Second, even if we were aware of such shifts, data necessary to adjust for them would be lacking.

The issues are radically different with respect to government owned, privately operated capital. Gordon's view that such capital should be included in private capital can be seriously questioned by reasoning along two lines. The first relates to the valuation of government output in the *OBE* national accounts, and the second to the pervasive reliance on the where-owned principle of classification in the construction of systems of national accounts.

III. *Capital Input in Production Function Analysis*

The first line of reasoning strongly suggests that for production function analysis of the whole economy, or the whole private economy, capital owned by government must be excluded from capital input if the national product and income estimates of *OBE* are used to measure output. The reason is that the value of the national product does not include payment for the services of government capital; and that the capital earnings component of the national income or

product, which is often used (e.g., by Edward Denison, Dale Jorgenson, Zvi Griliches, and Robert Solow) as weight for capital in calculating the contribution of capital to the national product does not include payments for the use of government capital.

From this vantage point it appears that the question is whether government owned, privately operated capital is like government capital for which the national product and income do not include a return, or whether it is like privately owned capital, for which a return is included. The answer to the question depends on pricing, particularly on the prices paid on government contracts, because most of the capital goods at issue are used chiefly to produce for the government. If the government pays the same prices to contractors who use government owned capital as to contractors who provide their own capital, and such contractors consequently earn the same dollar amounts as if they had invested their own funds, then government owned capital might be included in the private capital stock. But surely it is not the government's intent that it should pay prices that cover a contractor's nonexistent capital costs. Prices paid on government contracts are set in many ways. Little can be said unequivocally, but it would seem very odd to make the general assumption that the government pays for its own capital twice. If it does not, one must not add such capital to the private capital stock for production function analysis; to do so would distort the interpretation of changes in the national product.

It may be helpful to illustrate this verbal argument. The example is in terms of net rather than gross measures, but this is not essential. Assume that the economy consists of an independent shoe factory and of a government aided shoe factory. The former employs 800 units of labor and 200 units of capital to produce 1000 units of shoes for private use. The wages of labor are $800, the profits of capital are $200, and the value of shoes is $1000. The government aided shoe factory has the same physical endowment of labor and of capital and produces the same number of shoes. It also pays the same wages to its employees. But its capital is provided free of charge by the government; there are no profits, and the value of the shoes, which it sells to the government, is $800. These data are exhibited in column I of the following table.

TABLE 1—EXAMPLE TESTING MEASUREMENT OF CAPITAL INPUT

	Time period	
	I	II
Independent factory		
Physical units		
Labor	800	800
Capital	200	202
Shoes	1000	1002
Values		
Wages	800	800
Profits	200	202
Shoes	1000	1002
Government aided factory		
Physical units		
Labor	800	800
Capital	200	200
Shoes	1000	1000
Values		
Wages	800	800
Profits	0	0
Shoes	800	800
National income		
Wages	1600	1600
Profits	200	202
Total	1800	1802
National product		
Private shoes	1000	1002
Government shoes	800	800
Total	1800	1802
Factor input		
Labor	1600	1600
Capital		
Private	200	202
Government	200	200
Total	400	402

Now assume that the independent shoe factory increases its capital by 2. Its profits will increase by $2 to $202. And, using the assumptions commonly made in production function analysis, its output of shoes will increase by 2, to 202. (1000 shoes times the 1 percent increase in capital times the share of capital (20 percent) in total factor earnings.) Labor inputs and wages are assumed to remain unchanged, and the value of shoes

will increase to $1002. These magnitudes are shown in column II of the table. No change is assumed in the operations of the government aided shoe factory. Accordingly, the column II entries are identical to the column I entries for this factory.

The table also shows national income as the sum of wages and profits originating in the two factories, and the national product as the sum of the value of shoe production. Increases of $2 appear in profits, national income, shoes produced for private use, and national product. Inasmuch as there have been no price changes, the national product shown equals both current-dollar and constant-dollar (real) national product.

It is easy to see from the way the example is constructed that the $2 increase in national product is due to the increased input of capital. The same conclusion follows if production function analysis using only the privately owned capital is applied: To obtain the contribution of capital ($2) to the increase in output, we multiply base period profits ($200) by the increase in privately owned capital (1 percent). (See the data on factor inputs assembled in the last section of the table.) If instead we were to use total capital, including government owned capital, in the calculation, we would obtain $1 as the contribution of capital ($200 base period profits multiplied by 1/2 percent increase in total capital), and would be misled.

The above reasoning shows that Gordon is mistaken in creating the impression that the inclusion of government owned, privately operated capital is necessary and sufficient for valid production function analysis. I want to add, however, that in my opinion neither does it establish the opposite conclusion. For instance, if the example is varied to incorporate a change in government owned, privately operated capital, the results do not fall into place as neatly. It can be argued that the difficulties that are likely to be encountered stem from shortcomings in the measurement of the real volume of government shoes and that, in the absence of these shortcomings, our initial conclusion to exclude government owned, privately operated capital would be equally obvious. I have argued along such lines in the past in a similar case that did not involve government capital (see Jaszi).

I am citing this reference not because I am sure that I was right, but to show that we are dealing with very difficult problems. I have no solutions for them, but should like to offer four propositions. First, I am inclined to believe that the problems we encounter in fitting government capital into production function analysis are not specific to government capital, but have much in common with a broader set of problems long known to national economic accountants under the heading of "government services to business." Second, I suspect that, given the present nature of national product measures, inclusions of government capital in capital inputs would produce erratic results instead of leading to a systematic improvement in production function analysis. Third, I surmise that the proposition that fully satisfactory production function analysis would require comprehensive and realistic imputations of rates of return to government capital is formally correct, but that it is not practically helpful as a guide to further progress in this area of work. These three propositions I put forward with varying degrees of doubt. I am sure, however, of my fourth proposition: Gordon's characterization of the omission of government owned, privately operated capital from the private capital stock as involving "glaring errors" (see, p. 221 of his article) is injudicious and intemperate.

IV. *The "Where-Owned" vs. "Where-Used" Criteria*

The last major line of reasoning to which Gordon's theoretical framework is vulnerable is well known to national income accountants, who have been concerned with it for many years. It has to do with the problems of applying two rival criteria—where-owned and where-used—to the construction of systems of national accounts. These problems have been discussed most frequently in connection with the industrial allocation of rental income. Gordon seems to take the view that it is an obvious, accepted proposition, needing no further explanation or support, that capital should be recorded on a where-used rather than a where-owned basis.

I shall try to explain below that, unfortunately, matters are more complicated than he supposes.

But suppose for a moment that Gordon were right. Many types of public capital (most importantly roads, bridges, but other categories as well) are "used" by the private sector in exactly the same sense as the manufacturing capital with which his paper deals. If we recognize this, Gordon's results are placed in a better perspective. Instead of appearing to have discovered and eliminated an obvious error from *OBE's* national accounts, he is seen even on his own premises to have drawn attention to only a small segment of a vast, troublesome area in which a great deal of further theoretical and empirical work will need to be done before viable solutions are found.

It is more important, however, to draw attention to some overwhelming difficulties to which Gordon's where-used criterion would lead, especially if an attempt were made to apply it to industrial or sectoral divisions of the national accounts. The entire system of accounts—viewed as an interrelated set of production, appropriation, saving-investment, and balance sheet accounts—is based on records that are kept on a where-owned basis. A departure from the where-owned basis such as Gordon suggests would destroy the consistency of these accounts. In particular, data on investment and capital stocks would no longer be compatible with data on profits and other elements of value-added. In evaluating the extent of the havoc that would be created, one must keep in mind that leasing arrangements are widespread and increasingly important within the private sector of the economy.

A comprehensive, systematic restructuring of the accounts to convert them from a where-owned to a where-used basis does not seem feasible. If, in addition, we consider that in many types of analysis, for instance, those relating to the distribution of wealth, the where-owned principle of classification is the only eligible one, it becomes apparent that an alternative solution should be entertained. This is to treat the input of capital used by given industries but owned elsewhere as an intermediate product in production function analysis. In considering this alternative, it should be recognized that something very similar is now actually being done with respect to labor input. For instance, if an industry employs its own lawyers, they are counted as part of its labor input. If the legal services are provided by separate law firms, they are counted as intermediate products.

This example illustrates two points: The where-used criterion is much shakier than Gordon seems to realize; and lack of its implementation for the labor factor does not cause an outcry from the practitioners of production function analysis. Should we not pause and review the situation in this light, before we attempt to enthrone the will-o'-the-wisp where-used criterion for the capital factor?

REFERENCES

E. F. Denison, *Why Growth Rates Differ*, Washington 1967, pp. 135–37.

R. J. Gordon, "$45 Billion of U.S. Private Investment Has Been Mislaid," *Amer. Econ. Rev.*, June 1969, *59*, 221–38.

G. Jaszi, "The Conceptual Basis of the Accounts: a Re-examination," in Conference on Research in Income and Wealth, *A Critique of the U.S. Income and Product Accounts*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 22, Princeton 1958, pp. 71–72, 127, 144.

D. W. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Rev. Econ. Stud.*, July 1967, *34*, 249–83.

V. Perlo, "Capital Input-Output Ratios in Manufacturing," *Quart. Rev. Econ. Bus.*, Autumn 1968, *8*, 29–42.

R. M. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *34*, 312–20.

R. C. Wasson, J. C. Musgrave, and C. Harkins, "Alternative Estimates of Fixed Business Capital in the United States," *Surv. Curr. Bus.*, Apr. 1970, *50*, 18–36.

D. G. Wooden and R. C. Wasson, "Manufacturing Investment Since 1929," *Surv. Curr. Bus.*, Nov. 1956, *36*, 8–20.

$45 Billion of U.S. Private Investment Has Been Mislaid: Reply

By ROBERT J. GORDON*

The major purposes of my article were to call the attention of economists to the physical capital used by private firms but financed by the government, to attempt a measurement of the orders of magnitude involved over the 1940–66 period, and to induce the Office of Business Economics (*OBE*) to refine my methods and keep track of the "mislaid" capital in the future. I have succeeded in the last of these aims, since earlier this year the *OBE* for the first time published a historical time-series estimate of government owned, privately operated (*GOPO*) capital, an estimate which for the manufacturing sector is remarkably similar to my own earlier measurement attempt. Having agreed with me that the *GOPO* data are important enough for the *OBE* to begin keeping track of them, George Jaszi in his comment states that he does not understand how to integrate the new data into the *U.S.* national accounts, and moreover, he appears to doubt whether his and my data are of any use for production function analysis. This reply attempts to allay Jaszi's doubts and demonstrates that the relevant issues are neither so complex nor difficult as he suggests. The first four sections below correspond to Jaszi's subheadings, and two short additional paragraphs are added at the end on the issues of statistical accuracy and originality raised in Jaszi's two opening paragraphs.

I. *Gross Investment in Government Surplus Capital*

Jaszi disputes my first recommendation regarding production function studies of private output where the aim is to measure gross, undepreciated capital input. I claimed that the portion of the capital stock of private firms consisting of assets originally financed by the government but later transferred to private owners should be valued at its original gross value, not the marked-down net sales value measured in the *OBE* investment accounts. I argued (pp. 223–24) that the *OBE* estimates "seriously understate the cost of the assets, which have thus been included in *gross* capital formation at their *net* depreciated value. In the investment statistics they are completely merged with all new capital built in the year of sale, e.g., 1947 and for depreciation and retirement calculations are treated as if their age was the same as new 1947 assets."

Jaszi claims that my solution is not "open" because it cannot neatly be integrated into the "truncated form of the *OBE* system." But the whole point of my paper is that investment and capital estimates used in production function studies should not be confined by the manacles of the present *OBE* system. Instead, the *OBE* should establish a *second* set of capital and investment accounts which allocate capital assets by sector of use and which value all assets at original undepreciated cost. Economists and other users of capital statistics would then have a choice of the new accounts or the present accounts, which allocate investment by sector of ownership and value transfers of assets at the sales price. Whereas the present *OBE* investment concept is equal to saving both of the aggregate economy and each sector, the new concept of "investment-in-use" would differ for each sector from the present concept and hence would intentionally not be equal to saving.

Specifically, the *OBE* would begin by allocating each capital asset by sector of use. Corresponding to the new concept "capital-in-use" at time t in sector i (K_t^{Ui}) would be "net investment-in-use" ($I_t^{Ui}=K_t^{Ui}-K_{t-1}^{Ui}$), which in turn would equal installations of

* The author is assistant professor of economics at the University of Chicago and a research staff associate of the National Bureau of Economic Research. His research was financed by a grant to the National Bureau from the Sloan Foundation.

newly produced capital (H_t^{Ui}), plus net purchases from other sectors valued at original undepreciated cost (P_t^{Ui}); minus similarly valued retirements (R_t^{Ui}). If all government capital were included, for all sectors taken together net purchases would add to zero ($\sum_i P^{Ui}=0$), so "gross investment-in-use" for the aggregate economy ($H_t^U=\sum_i H_t^{Ui}$) would be equal to the final sales of the capital goods industry (adjusted for net exports). Total *private* gross investment in use would differ from the present *OBE* concept of gross private domestic investment by the amount of investment in *GOPO* and the revaluation of assets transferred between the government and private sectors.

In short, there is no reason for the *OBE* to limit itself to compiling investment and capital data by a single concept which may not be appropriate for all purposes. The problems raised by Jaszi's first section disappear when we recognize that the new system would supplement rather than replace the present system. Economists studying portfolio allocation, the demand for investment goods, and the effects of wealth on spending decisions would prefer to utilize the present *OBE* investment and capital data allocated by sector of ownership, but in addition, many students of technology and productivity would prefer a separate set of statistics allocated by sector of use even if this would break the link between saving and the new concept of investment.

The need for a dual set of concepts would be greatest in a hypothetical economy in which all manufacturing plants and equipment were owned by one sector, say banks, and leased to manufacturers. A single system of statistics allocating capital by sector of ownership, as at present in the United States, would be of little use in studying the level and rates of change of manufacturing capital-labor and capital-output ratios of the hypothetical economy. In the *U.S.* economy there is a considerable overlap between sector of ownership and sector of use, of course, but in some cases, as with the *GOPO* capital measured in my paper, ownership and use do not coincide even for the aggregate private sector.

I am mystified how Jaszi could have read my article as suggesting that the present *OBE* concepts of saving and investment should be *replaced* rather than supplemented. The point of my article was simply that an important component of capital used by *U.S.* private firms was at that time unmeasured and was therefore overlooked in production function studies: I argued on p. 222 that "... the *U.S.* national accounts have failed to keep track of government capital. Government purchases of industrial structures and equipment are lumped in with all other government current purchases of goods and services, and there is no published series on the amount of government capital which has been used during the past 25 years to produce private output."

II. *Gross Stocks of Government Surplus Capital*

Jaszi's second point is that my proposal to value surplus government capital sold to private purchasers at its original cost rather than sales value may overstate its contribution to private product. He is very hesitant in making this claim, since my "proposal has a good deal of plausibility." His only argument against my suggestion is that the surplus capital was the part of investment required by the war that "business was unwilling to finance privately ... the investment judged least likely to succeed in a civilian reincarnation." Yet the process of expanding America's industrial capacity during World War II did not involve the kind of careful marginal judgments by businessmen which Jaszi assumes to have taken place. The government did not ask businessmen whether they were interested in financing aircraft plants; the government simply assumed the burden of financing without question.[1] And, when the war was over, specialized machinery was scrapped, but in these industries a substantial portion of the wartime investment consisted of general purpose factory buildings and general purpose machine tools (e.g. lathes) which were as useful

[1] "In the war-oriented explosives, ammunition, ordnance, aircraft, and ships categories, for instance, the Government financed 93 percent of facilities expansion and private firms the remainder" (Gordon, p. 225).

in civilian production as privately financed plant and equipment of the same vintage. As was pointed out in my article, the sales price of this surplus plant and equipment substantially underestimated its likely profitability in peacetime production:

> The returns received . . . on plants sold after the war were bound to be below original cost because of depreciation, of course, but a more important cause of low realizations . . . was undoubtedly a general lack of certainty that wartime prosperity could be maintained. Businessmen were hesitant to commit large sums on war-built plants which they feared might be badly underutilized as peace broke out. Thus many firms acquired modern, well-equipped plants in 1946 at bargain prices which seriously understate their full-capacity ability to produce. Part of this plant capacity may have gone unutilized at first, but utilization doubtless improved in later years as private demand and output increased far above expectations. [p. 234]

III. *Capital Input in Production Function Analysis*

Jaszi's numerical example is designed to argue that if the contribution of *GOPO* capital to private output is not included in national income, then *GOPO* capital should not be included in factor input in production function analysis. The numerical example is unnecessary, for Jaszi is wrong on two counts: 1) prices paid on government contracts *do* allow private contractors to make a return on government-owned capital—a return included in private output—even though this procedure does, as Jaszi points out, "seem very odd"; and 2) because the equality of capital's income share and capital's contribution to output is an assumption, not a proved fact, and one of the aims of production function analysis is to investigate the validity of this claim.

If private contractors earn a return on government owned capital, that capital should be included in the private capital stock (for the purposes of production function analysis and, not, of course, for calculations of private wealth). Jaszi recognizes this, but denies that such a return is earned by private contractors, since this would imply that "the government pays for its own capital twice." There is no way of answering Jaszi's question whether the government pays the same prices to contractors who use government owned capital as to contractors who provide their own capital. The question implies that products exist which are made by both types of contractors, a situation which does not occur. Rather, *GOPO* capital is used to make aircraft, plutonium, and other unique products which are not simultaneously being manufactured by contractors using privately financed capital. There is no industry in which Jaszi's "independent" and "government aided" factory exist side by side, both making the same product.

Although we cannot provide empirical evidence on Jaszi's nonexistent two-factory situation, it is easy to document that private contractors using *GOPO* capital are paid substantial fees which are included in national income and which are recognized by the government to be profits. These fees are paid in addition to reimburseable costs for labor, materials, and management services under the standard "cost plus fixed fee" (CPFF) contract system. In the atomic energy industry, where almost all capital falls into the *GOPO* category, ". . . *CPFF* contracts constitute almost all of the government-business relationships. Of the Atomic Energy Commission's expenditures 95 percent is made through contractors, and 85 percent of the value of these expenditures is in *CPFF* contracts" (Richard Tybout, p. 10). Regarding *GOPO* defense plants other than atomic energy, "Despite much criticism of the *CPFF* contract in general and many attempts to eliminate it, the use of this arrangement in the case of government owned contractor-operated plants was never seriously questioned" (J. P. Miller, p. 129). "The *CPFF* contract became the principal instrument for procuring novel or expensive items such as aircraft, heavy ordnance equipment, and ammunition. Before the end of World War II, the Army *CPFF* commitments would exceed $50 billion" (Richard G. Hewlett and Oscar Anderson Jr., p. 187).

There is little doubt that the fees paid on

CPFF contracts represent pure profit to the private firms involved. Discussions of contracting procudures often slip from calling these payments "fees" to the more accurate usage "profits": ". . . the more novel, complex, and difficult the work, the more profits to which the contractor is entitled. Insofar as atomic energy work is novel, there are uncertainties to be resolved and hence work to be done of the type for which profits are compensation" (Tybout, p. 62). One justification for these profit payments seems to be that if the private contractor were not working on the government contract, it could be making greater profits elsewhere: "We believe that private concerns rate some compensation for the profit they are not able to get by engaging those key people in their own line of operations" (Carleton Shugg, pp. 2157–58). This argument, of course, neglects the fact that unless the contractor's privately owned capital is underutilized, the "key people" in question would not be able to earn profits for the company in alternative pursuits unless the company were to invest in additional privately financed capital for them to work with. Thus these profit payments are compensation for government-financed capital, however little justification can be made for them.

A detailed accounting of the *CPFF* fees is beyond the scope of this reply, but in the atomic energy industry most contractors appear to be paid fees in the range of 2 to 6 percent of estimated operating expenses (Tybout, p. 66). For instance, in the 1949–53 period, the fees of the Carbide and Carbon Chemicals Company were 2.12 percent of expenses, and of Monsanto were 5.56 percent of expenses (Tybout, pp. 68–69). The practices of the Atomic Energy Commission (*AEC*) appear to be relatively conservative compared to the other major owners of *GOPO* capital, the Defense Department (*DOD*) and the National Aeronautics and Space Administration (*NASA*): ". . . But *AEC* fees have remained lower than those of the *DOD* and *NASA*, particularly in large contracts, for they have followed curves in which the fee declines proportionately more, as the size of the contract rises, than *DOD* and *NASA* fees. Industry has sought to raise them, and, in general to get the *AEC* to adopt the more generous *DOD* policies . . ." (Harold Orlans, p. 199). Moreover, it is not the size of the fees relative to private returns in other industries that is important here but the *existence* of the fees as a return on government owned capital paid to private firms. Jaszi would surely agree that we should include the plant and equipment of the Penn Central Railroad as part of the *U.S.* private capital stock even though in recent years its rate of return has been low and sometimes negative.

A less obvious return on government capital paid to private firms, in addition to the fees themselves, is the increase in profits on privately financed capital resulting when private firms exploit knowledge and discoveries made in government financed research with which these companies are associated. "A company may spend, say 1 or 2 percent of its gross income on its own research and development work; but when they do government research and development they thereby get large additional sums of money to do such work. In this way they enhance their competitive position without having to use their own money" (Hyman G. Rickover, p. 3). "There can be little doubt that some firms have obtained from their government contracts a commerical advantage in their private nuclear business. It is no coincidence that the two monarchs of the civilian nuclear power business, Westinghouse and General Electric, have long operated major *AEC* reactor laboratories" (Orlans, p. 20).

The second main argument in favor of including *GOPO* capital in the private capital stock for production function analysis, in addition to the profits made on *CPFF* contracts, is the possible invalidity of the standard competitive assumption that factors are paid their marginal products in all sectors. Assume the following simplified economy: there are two sectors, agriculture and manufacturing, and labor is not paid its marginal product in agriculture, where the wage is positive despite the fact that one-third of the agricultural labor force is redundant and hence labor's marginal product is zero. While farm capital is privately owned, all manufacturing capital is owned by the government and none of the return to capital is included

TABLE 1—A DUAL ECONOMY WITH GOVERNMENT OWNED CAPITAL

	Time Period					
	I			II		
	Agric.	Mfg.	Total	Agric.	Mfg.	Total
Labor Force	150.0	50.0	200.0	100.0	100.0	200.0
Capital-in-Use	100.0	100.0	200.0	100.0	200.0	300.0
Total Product[a]	100.0	70.7	170.7	100.0	141.4	241.4
National Income[b]	100.0	35.4	135.4	100.0	70.7	170.7
OBE Capital	100.0	0.0	100.0	100.0	0.0	100.0
OBE Capital/National Income	1.00	0.00	0.74	1.00	0.00	0.59
Capital-in-Use/Income	1.00	2.82	1.48	1.00	2.82	1.76

Notes:

[a] The production functions in each sector are Cobb-Douglas with no technical change, exponents of .5 on both labor and capital, and with 50 redundant agricultural workers excluded from effective labor input in period I.

[b] Income from manufacturing capital is excluded from national income. Factor shares are equal to Cobb-Douglas exponents.

in national product. To make the calculations simple, we assume a Cobb-Douglas production function in each sector, with labor input in agriculture only counted as the two-thirds who are productive.

To make the point at issue, let the capital stock in manufacturing double and let this be sufficient to draw all of the redundant agricultural workers into manufacturing. In the numerical example shown in Table 1, national income increases by 26 percent even though the total population and the *OBE* concept of privately owned capital remain constant. A mysterious decline in the ratio of *OBE* capital to output has occurred, with no apparent change in factor input. A supplementary capital series which includes *GOPO* capital would, unlike the *OBE* series, provide a clue as to the source of the decline in the *OBE* capital-output ratio: a shift in the composition of output occurs simultaneously with an increase in capital-in-use. This example is not far fetched, since a portion of the extraordinary decline in the *U.S.* ratio of *OBE* capital to output between 1929 and 1948 (which motivated my article) may have been due to the increased productivity of previously redundant workers made possible by government financed plants and equipment.

IV. *The "Where-Owned" vs. "Where-Used" Criteria*

Jaszi concludes by characterizing the where-owned and where-used criteria as rivals, requiring that one or the other be chosen as the sole basis for setting up the accounts. The where-used criterion would, as Jaszi rightly says, destroy the consistency of the present system of accounts. But nowhere in my article did I suggest scrapping the present system of accounts; I merely recommended supplementing them with where-used data suitable for production analysis. This point is the same as that made above in Section I.

In addition, Jaszi uses the examples of roads and bridges to claim that my article draws attention to only a small portion of the capital financed by the government but used by the private sector. Yet this point is made explicitly on page 237 in my article: "In addition to the assets discussed above, for instance, much of the apparent decline in the *U.S.* private capital-output ratio may have been due to the gradual replacement of privately owned railroad assets by government owned highways and airports."

V. *Statistical Techniques.*

In his first paragraph Jaszi alludes to

"serious flaws" in my estimates discovered at the *OBE*. I am delighted to have Jaszi's talented experts revise and refine my estimates, since one of my main purposes was to induce them to do so. Thus far, their major discovery has been the vast government investment in privately operated merchant ships, which I neglected to include in my $45 billion. But for the manufacturing sector on which my discussion concentrated, the new *OBE* estimates of gross capital stocks are remarkably close to mine. The figures (in billions of dollars) are as follows, with my figures given first: 1945, 30.9 vs. 30.8; 1950, 24.3 vs. 26.3; 1955, 29.3 vs. 27.9; 1960, 26.7 vs. 25.9; and 1965, 25.8 vs. 23.2.[2]

VI. *Originality*

Jaszi's second paragraph attempts to read into my article a claim of original discovery which it does not contain. In fact, Jaszi will look in vain for the word "discovery." My claim was not that no one at *OBE* knew about the existence of *GOPO* capital but that 1) the economics profession and particularly production function investigators had remained ignorant of *GOPO* capital, 2) that no one had previously constructed a time-series of *GOPO* capital suitable for use in production function studies, and 3) that no one had previously published a discussion of measurement problems encountered in estimating *GOPO* capital. Regarding 1), I described *GOPO* capital accurately as "little-known" and "neglected assets," and the absence of any discussion of them in production function studies during 1957–69 indicates that Wooden and Wasson's 1956 article (to which Jaszi refers) was ignored or overlooked.[3] Regarding 2), neither the 1956 article nor any other *OBE* publication before 1970 presented a historical time-series of the capital concepts in question. Regarding 2) and 3), the article by Victor Perlo cited by Jaszi was published six months after my manuscript was submitted to this *Review* and eighteen months after my doctoral thesis containing similar material was presented to Massachusetts Institute of Technology. Further, Perlo admits his estimate is "a rough, minimal adjustment" which is described in a single sentence (1968, p. 38). Finally, in a recent article, Perlo cites my article rather than his own as having "found" the *GOPO* capital (1970, p. 24).

[2] *OBE* estimates are from Wasson et. al., Table 4, col. 5, minus Table 1, col. 5, plus Table 7, col. 5. My estimates are calculated from Table 4, p. 233 of my article, for manufacturing only.

[3] This is not surprising, since the 1956 reference in question contains only one column on *GOPO* assets buried in the middle of a 26-column article (Wooden and Wasson, p. 14).

REFERENCES

R. J. Gordon, "$45 Billion of U.S. Private Investment Has Been Mislaid," *Amer. Econ. Rev.*, June 1969, *59*, 221–38.

R. G. Hewlett and O. E. Anderson, Jr., *The New World, 1939–1946,* University Park, 1962.

J. P. Miller, *Pricing of Military Procurements,* New Haven 1949.

H. Orlans, *Contracting for Atoms*, Washington 1967.

V. Perlo, "Capital Input-Output Ratios in Manufacturing," *Quart. Rev. Econ. Bus.*, autumn 1968, *8*, 29–42.

———, "Arms Profiteering: It's Not a Myth," *The New Republic*, Feb. 7, 1970, 23–5.

H. G. Rickover, U.S. Senate, Committee on Small Business, Subcommittee on Monopoly, *Patent Policies of Government Departments and Agencies—1960,* 86th Cong., 2d sess., Washington 1960.

C. Shugg, U.S. House of Representatives, Committee on Appropriations, *Independent Offices Appropriations for 1951,* Hearings 81st Cong., 2d sess., Washington 1950.

R. A. Tybout, *Government Contracting in Atomic Energy,* Ann Arbor 1956.

R. C. Wasson, J. C. Musgrave, and C. Harkins, "Alternative Estimates of Fixed Business Capital in the United States," *Surv. Curr. Bus.*, Apr. 1970, *50,* 18–36.

D. G. Wooden and R. C. Wasson, "Manufacturing Investment Since 1929 in Relation to Employment, Output, and Income," *Surv. Curr. Bus.*, Nov. 1956. *36*. 8–20.

Liability in Law and Economics: Note

By Thomas E. Borcherding*

In his article on negligence and liability, in this *Review*, Simon Rottenberg addresses himself to the question: "What is the appropriate rule for the compensation of persons damaged by accident?" (p. 107). He quite correctly asserts that the ideal rule does not necessarily minimize the cost of damages resulting from negligence, but rather structures the decision-making process in such a way as to result in the optimal amount of harm. This ideal state is reached when the marginal social value of resources devoted to accident prevention is just equal to the marginal social benefits of increased safety.

Rottenberg illustrates this principle by examining the current law of liability in the international air-carrier industry. These rules, stated in the Warsaw Convention of 1929 and the Hague Protocol of 1955, set the liability per international air passenger at a maximum of $16,000. Offering evidence that the present value of life and limb of the representative passenger exceeds this figure, he concludes that the current law of liability " . . . promotes an insufficient attention to care, measured by the quantum of resources used to prevent accidents" (p. 112).[1]

As a superior alternative, Rottenberg suggests the adoption of a rule of unlimited liability.[2] This would not, he states, prevent all accidents, but by placing the full cost of death or injury on the airline it would lead to a more satisfactory allocation. In terms of the standard welfare criteria requiring marginal equivalences: " . . . rational maximizing airlines will find a margin at which it will pay to permit accidents to occur rather than incurring the cost of preventing any more of them. . . . The social welfare will be maximized by such behavior" (p. 112). Though this last statement is not strictly incorrect, I find it wholly misleading. Following Ronald Coase's theorem, I will show that the quantity of resources devoted to safety is unaffected by the locus of liability.[3]

To demonstrate my point let us analyze the behavior of a firm providing air service under various liability rules. Purely for purposes of simplification we will assume that this firm is imbedded in a competitive industry, that it flies one given route, and that it uses one type of aircraft seating n passengers and traveling under unvarying flight conditions. The first assumption assures us that the firm sets prices nondiscriminately and at marginal cost. The remaining assumptions provide a convenient way of holding the transportation component of air service, T, constant so that safety, S, can be scrutinized alone.

Safety is defined as a unit of services yielding a certain probability, p, of successfully completing a journey without harm to either life or limb; i.e. $p=p(S)$ and $0 \geq p \geq 1$. Not only must $p'(S)$ be positive but $p''(S)$ must equal zero to insure that all safety units are of the same quality.

The total cost of the air service will be $C(\overline{T}, S)$. Under competitive conditions, the firm's output of safety will be set where the price of the marginal unit of safety equals its marginal cost and the rents are just sufficient to cover the transport cost. Since

* Assistant professor of economics, University of Washington. I wish to thank John McGee, Yoram Barzel, and Allan Hynes for their suggestions.

[1] Actually, since Rottenberg's article the limit of liability has been raised to $75,000 for passengers whose flights originate in, are destined for, or are via the United States. For all others, the former rules hold. But even $75,000 is far less than the present value of the life of the average international air traveler. Thus, we are still analyzing the effects of a limited liability rule.

[2] Rottenberg entirely neglects the welfare distribution effects. Implicitly, this amounts to assuming the welfare superiority of any Paretian allocation over any non-Paretian one. This would also seeem to imply the welfare indifference of one Paretian state to that of another. These assumptions are conspicuously common among applied welfare theorists.

[3] I assume here that transactions costs are invariant to the rule chosen and that demand and supply functions are unaffected by the different income distributions that result under each rule.

in this analysis transportation is a parameter, nothing is lost if it is dropped as an argument in all the total and marginal conditions that follow. Let us now analyze the decision with respect to safety under conditions of zero, of unlimited, and of the current rules of liability.

First, consider the zero-liability rule. Let $C'(S)$ be the marginal cost of safety provision and $M_i'(S)$ be the marginal evaluation schedule for each of the n passengers. At equilibrium all the $M_i'(S)$ functions will be identical, since the firm is constrained to act competitively (price discrimination is not feasible) and since safety is a jointly consumed good for each of the n passengers. Assuming numerous alternatives, any passenger will react to disequilibrium by finding a price-safety package more to his liking.

Under zero liability, the full trading equilibrium that holds is

$$(1) \qquad \sum_{i=1}^{n} M_i'(S) = G'(S) = C'(S)$$

$G'(S)$ is the group's marginal evaluation and equals $nM_i'(S)$. Second-best considerations aside, this allocation is Paretian,[4] since the marginal social benefit of safety (the sum of individual marginal evaluations) equals the marginal social cost.[5]

Now contrast this zero-liability situation with Rottenberg's allegedly superior unlimited liability rule. Under such a law the level of awards expected for any level of safety, $A(S)$, would depend on the total awards actually made to the group in the event of an accident, B, multiplied by the probability of an accident given that level of safety; i.e., $A(S) = [1-p(S)]\,B$. Marginal awards will equal the change in that probability times the actual awards; i.e., $A'(S) = -p'B$. Since these amount to a marginal reduction in awards, passengers will treat them as a marginal loss and carriers will view them as a marginal benefit. Since homogeneity of the quality of safety units requires that p'' be zero, marginal awards are constant with respect to output.

Under a full-liability regime, equilibrium requires that

$$(2) \qquad \sum_{i=1}^{n} \hat{M}_i = \hat{G}_i'(S) = C'(S)$$

The group's net marginal evaluation of safety, $\hat{G}'(S)$, equals the direct marginal value, $G'(S)$, plus the passenger's expected marginal loss of awards, $A_e'(S)$; $\hat{C}'(S)$ is the firm's net marginal cost of safety provision and equals the direct marginal cost, $C'(S)$, plus the expected marginal fall in awards, $A'(S)$. Assuming adequate information, firm and passenger estimations of $A'(S)$ will coincide and (2) can be rewritten:

$$(2') \qquad G'(S) + A'(S) = C'(S) + A'(S)$$

This is identical to (1), except for distribution effects.

Under a partial-liability rule (like the Warsaw-Hague agreements), the expected payment for injury or death will be lower than under unlimited liability; therefore $A'_f(S)$ will also be less. However, aside from distributional consequences it, too, is the same as (1).

All of this was predictable from the start. Since no external effects are assumed, alteration in liability rules introduces no new cost or benefit into the safety decision but merely shift costs between contracting parties. Equilibrium will hardly be affected by marginal effects that are mutually canceling.

[4] Since all passengers on a given flight consume the same bundle of T and S, we know that $T_i = T$ and $S_i = S$ for all passengers. The total value placed upon a given bundle $G(S)$ is $\sum_{i=1}^{n} M_i(S)$ and the total cost is $C(S)$. Ignoring second-best considerations, Pareto optimality requires:

$$W_{\max} = \sum_{i=1}^{n} M_i(S) - C(S),$$

where

$$(\text{i}) \qquad W' = \sum_{i=1}^{n} M_i'(S) - C'(S) = 0$$

and

$$(\text{ii}) \qquad W'' = \sum_{i=1}^{n} M_i''(S) - C''(S) < 0$$

Competition insures that both (i) and (ii) will hold.

[5] I am neglecting, as does Rottenberg, any third-party damages to nonpassengers and noncarriers, such as may occur when planes crash in cities and suburbs.

This result also holds for the monopolistic organization of the industry. Assuming simple monopoly (nondiscriminatory) pricing and numerous alternatives (each passenger is in full consumer equilibrium), the invariance of equilibrium to liability rules can again be easily demonstrated.

Under zero liability, the firm maximizes the function

$$\pi = SG'(S) - C(S)$$

Assuming the sufficient condition holds, the marginal condition becomes

$$(3) \quad \pi' = G'(S) + SG''(S) - C'(S) = 0$$

This is, of course, the familiar rule that marginal revenue equals marginal cost.

With unlimited liability, the firm maximizes

$$\begin{aligned}\hat{\pi} &= S\hat{G}'(S) - \hat{C}(S) \\ &= S[G'(S) + A'(S)] - [C(S) + A(S)]\end{aligned}$$

The necessary conditions here are

$$(4) \quad \begin{aligned}\hat{\pi}' &= G'(S) + A'(S) + SG''(S) \\ &\quad + SA''(S) - C'(S) - A'(S) = 0\end{aligned}$$

Since $A''(S)$ is necessarily zero, (4) becomes identical to (3). Partial-liability rules will yield the same answer. In each case, S is identical in quantity and is suboptimal. Under perfectly discriminating monopoly $SG''(S)$ is zero; hence, for any rule of liability the allocation is ideal as in (1) or (2′), differing again only in distributional consequences. In the monopoly situation where numerous alternatives are not available to consumers and perfect discrimination is not feasible, not all consumers will be in equilibrium. I am unable to posit formal equilibrium conditions for each (or any) rule of liability in such a situation, but ignoring income effects, I would conjecture that because of the mutually offsetting nature of awards at the margins, allocation will be unaffected by the rule chosen. In short, it appears that for any market structure the effect of liability law is purely distributive.

One caveat ought to be made, however. If passengers find it difficult to ascertain the quantity and/or quality of the safety services provided by carriers, they are faced with a risky and uncertain set of variables from which choice must be made. In such situations passengers would have uncertain total evaluations for safety and, hence, the marginal value of safety to them would not be readily ascertainable.[6] In a zero-liability regime markets would doubtless develop to aid rational choice by consumers,[7] but such devices are costly to institute and are subject to error as well. On the other hand, in a world of full liability the courts would give carriers crude signals as to the marginal value of safety.[8] This process would also be costly in terms both of transactions and of error. It is probably true that in the real world the sum of the costs of transactions, of the loss from erroneous choices, and of the cost of securing information would differ under any liability rule. I cannot even guess whether upon closer inspection the empirical evidence would sustain Rottenberg's desideratum. It is, however, an important and intriguing challenge.

REFERENCES

R. H. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

S. Rottenberg, "Liability in Law and Economics," *Amer. Econ. Rev.*, Mar. 1965, *55*, 107–14.

[6] If consumers grossly overestimate the quantity of safety provided as compared to producers' estimates, a shift from zero to partial or full liability might be advisable. If consumers underestimate this flow, then a move towards zero liability might improve allocation. One would expect, however, that if such gross error were present market forces would emerge to provide better information. These are briefly discussed in fnn. 7 and 8.

[7] Insurance brokers would certainly provide information in the form of differential rates with respect to carriers. In addition to aiding rational consumer choice, these rates would help the market to establish competitive outcomes.

[8] Under full liability the awards set by courts will not lead to zero net marginal evaluations $[G'(S) = A'(S)]$ for all consumers. Those preferring risk might actually be willing to sell back some of their expected claims for cash and less safety. Risk-averse individuals would be willing to buy more than is made available on the basis of court signals alone. All this presumes, however, that the cost of such adjustments is low relative to the gains realized.

On Navigating the Navigation Acts with Peter D. McClelland: Comment

By JOSEPH D. REID, JR.*

Peter McClelland recently joined Robert Thomas in rejecting Lawrence Harper's earlier finding that the British Navigation Acts placed "... a heavy burden upon the colonies."[1] While I agree with Thomas and McClelland's conclusion that the cost to the colonists of the trade route distortions attributable to the Navigation Acts was probably small, I believe the hitherto unpresented correct analytical approach to estimating the costs of these trade distortions is of sufficiently general applicability in empirical investigations bedeviled by "missing" data to merit illustration by navigating the Navigation Acts one more time.

The Navigation Acts "... were a series of laws designed to alter the trade of the British Empire in the interests of the mother country" (Thomas, p. 619) by: 1) regulating the manning and ownership of vessels carrying trade; 2) constraining the potential geographical sources of import supply and export demand; 3) encouraging specific industries by preferences and bounties. Thomas' assessment attempts to estimate costs to the colonists of the Navigation Acts under all three facets, leading Thomas to estimate a *net* cost therefrom of \$2,660,000 or \$1.20 per capita in 1770 and an average *net* cost of \$2,255,000 or \$1.24 per capita over the period 1763–72 (Thomas, Table 2, p. 637), or approximately 1.2 percent of estimated national income. McClelland only considers the trade distortions arising from the requirement that certain enumerated imports to and exports from outside the Empire pass through England. He directly estimates that 3 percent of national income "... is an upperbound measure of the actual costs involved" (McClelland, p. 376).

But Thomas seems to have and McClelland clearly calculated the costs to the colonists of the London entrepôt requirement on goods enumerated under the Navigation Acts as approximations of the area $P_{C1}P_{C0}BA$ in Figure 1c, where S_{XA} represents the f.o.b. the colonies (excess) general equilibrium supply schedule of colonial exports affected by the Navigation Acts to areas outside the British Empire—represented as reexports to Amsterdam by Thomas and McClelland; D_A represents the outside the Empire (Amsterdam) general equilibrium demand schedule less direct freight for enumerated colonial exports, i.e., effective demand in the absence of British constraints by outside the Empire demanders at the colonial ports of exit; P_{C1} represents the *actual* pre-Revolution f.o.b. price of the export in question at its chief colonial port of exit; D'_A represents the actual effective demand schedule by outside the Empire demanders; and P_{C0} represents the counterfactual would-have-been-received price at its chief exit port in the absence of the Navigation Acts.[2] Hence, the Acts' extrepôt requirements are analytically equivalent to an (admittedly inefficient) excise tax on enumerated trade out of the Empire, with a cost to the colonial equal to the effect upon price times the affected quantity. However, Figure 1 indicates $P_{C1}P_{C0}BA$ is the whole cost to the colonies only if effective English demand for the export in question is zero—i.e., only if the non-zero

* The author wrote this note while a National Science Foundation Graduate Fellow at the University of Chicago. He is indebted to Robert W. Fogel, the managing editor, and the referee for many helpful suggestions.

[1] Harper (1942, p. 4, also 1939), McClelland, and Thomas. As J. R. T. Hughes points out in his discussion of McClelland, what Thomas adjudged *insignificant* Harper labelled *heavy* (see Hughes, p. 382).

[2] The schedules incorporate the effects of changes in the terms of trade and in other prices consistent with general equilibrium at the indicated price and quantity in the illustrated export market.

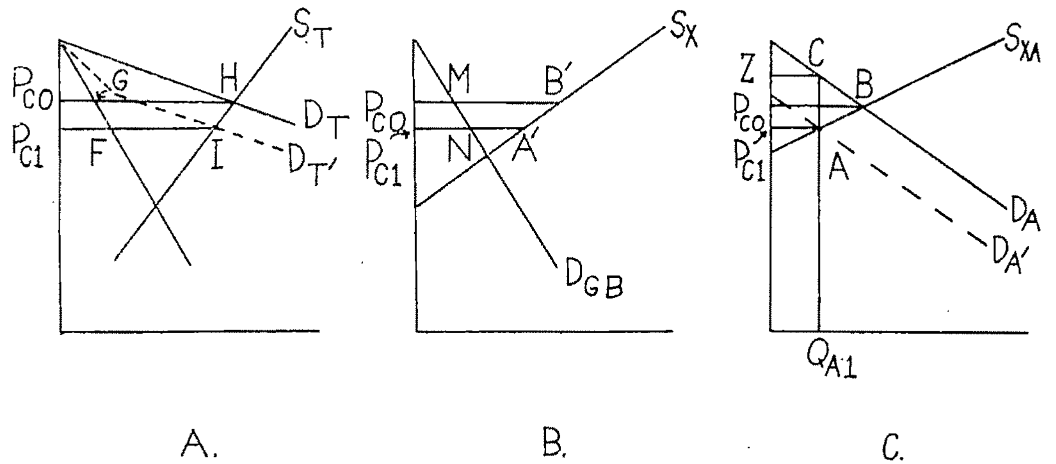


FIGURE 1

demand points on the English demand schedule all occur at a price less than P_{C1}. Clearly the true total cost on colonial exports of the London entrepôt provisions of the Navigation Acts is represented by the area $P_{C1}P_{C0}B'A'$ in Figure 1b—where S_X represents the (excess) supply of the colonial export in question to both within the Empire and without the Empire demanders—or the equal area $FGHI$ in Figure 1a—where S_T represents the total colonial supply schedule of the export in question. The Navigation Acts English entrepôt requirement was equivalent to a downward shift by the amount of additional freight and handling costs of the effective outside the Empire f.o.b. demand for enumerated colonial exports. This outside the Empire fall in effective f.o.b. demand occasioned a decline in f.o.b. prices of colonial exports for outside the Empire—thus leading to the colonial loss $P_{C1}P_{C0}BA$ on reexports estimated by Thomas and McClelland (Figure 1c)—and a redirection of previous production for outside the Empire consumption to English and domestic sale, leading to price declines and attendant loss to the colonists indicated by the area $P_{C1}P_{C0}MN$ on within the Empire exports (Figure 1b) and the transfer from colonial producers' surplus to colonial consumers' surplus of the amount indicated by $P_{C1}P_{C0}GF$ in Figure 1a.

In making his earlier estimates, Harper explicitly recognized the cost to the colonists of the entrepôt provisions of the Navigation Acts was the sum of the cost on outside the Empire (Amsterdam) and within the Empire (London) colonial exports. Concerned with estimating a lower bound of the cost, Harper consciously ignored the entrepôt provision effects upon London consumed exports and London originating imports.[3] Since Thomas and McClelland stress the smallness of the Acts' burden upon the colonists, their failure to incorporate the Acts' effects upon London prices and quantities clouds the "upperness" of their upper bound estimates and needs further justification.[4]

Within the (too narrow) boundaries of Figure 1c, McClelland's criticisms of Harper's estimates of $P_{C1}P_{C0}BA$ are justified: Harper's period is too long and the required price estimates are too diverse to make a general deflation of post-Revolutionary War prices at Philadelphia (as measures of P_{C0}) acceptable; and Harper's two alternative estimates—contemporary testimony and recorded value of exports minus value of imports—are based upon poor data and

[3] "Yet for conservatism's sake it may be well to assume that it [the Navigation Act cost] was borne only by reexported tobacco" (Harper 1939, p. 35).

[4] Especially since McClelland estimates that between 10 and 50 percent of enumerated transatlantic exports were not reexported (p. 377).

imply that the Amsterdam demand schedules for colonial exports were of infinite elasticity.[5]

But McClelland's criticisms of Thomas' approximation of the colonists' loss $P_{C1}P_{C0}BA$ are less than insightful and don't properly illustrate the novelty and ease with which Thomas is able to indirectly estimate the "missing" counterfactual price, P_{C0}.[6] Thomas and McClelland wish to estimate the loss on enumerated colonial exports to Amsterdam stemming from the Navigation Acts requirement that these exports pass through London: in Figure 1c the area $P_{C1}P_{C0}BA$. Now, the actual f.o.b. price of colonial exports, P_{C1}, is known, as is the amount reexported from England for outside the Empire consumption. Similarly, Amsterdam prices of colonial exports are available for the pre-Revolution period. Thus P_{C1} and $P_{C1}A = Q_{A1}$ are known for affected exports. To estimate the trapezoid in Figure 1c, McCelland follows Harper in examining contemporary testimony on tobacco shipping costs, then examines England to Northern Europe additional shipping costs for tobacco during 1712–18 (about 12-1/2 percent of transatlantic rates), sugar in the 1730's (16 percent) and rice in the 1760's (18 to 30 percent) to posit that the proportional increase in colonial freight rates occasioned by the Navigation Acts ". . . between 1763 and 1775 was no greater than 40 percent" (McClelland, p. 380). Justifying his following of Harper's previously criticized footsteps by distinguishing his search for an *upper* bound from Harper's search for a lower bound for this burden, McClelland then assumes (implicitly) that the entrepôt requirements of the Acts occasioned no change in the Amsterdam colonies terms of trade and estimates an upper bound $P_{C1}ZCA$ as an estimate of $P_{C1}P_{C0}BA$ in Figure 1c. McClelland argues $P_{C1}ZCA$ will be larger than (and therefore an overestimate of) $P_{C1}P_{C0}BA$ as long as the supply curve does not have ". . . an extreme slope relative to the slope of the demand curve."[7] McClelland's laborious direct estimation of $P_{C1}P_{C0}BA$ utilizes sketchy and ambiguous data to derive a probable overestimate (by an uncertain magnitude) of the desired amount, even ignoring that the desired amount is only part of the true cost.

Thomas initiates his attempt to estimate $P_{C1}P_{C0}BA$ in Figure 1c by estimating the in-the-absence of Navigation Acts price P_{C0} ". . . by dividing the observed Amsterdam price . . . before the Revolution by the ratio of Amsterdam to Philadelphia . . . prices after the Revolution" (p. 624); i.e., Thomas' estimate of P_{C0} is

$$\hat{P}_{C0} = \frac{P_{A1}}{P_{A2}} P_{C2} \tag{1}$$

where

P_{A1} = the Amsterdam price of the enumerated export before the American Revolution

P_{A2} = the Amsterdam price of the enumerated export after the Revolution

P_{C2} = the f.o.b. America price of the enumerated export after the Revolution

To interpret Thomas' estimate, let the demand and supply schedules for colonial exports be approximated within the relevant ranges as

$Q_D^A = DP_A^{-d}$, the Amsterdam demand schedule for the export in question; D is a shift term, P_A is the Amsterdam price, d is the (positive) elasticity of demand

$Q_S^C = SP_C^s$, the colonial f.o.b. supply schedule; S is a shift term, P_C is the colonial price f.o.b., s is the elasticity of supply

fP_C = the direct cost per unit freight from the colonies to Amsterdam

tP_C = the total additional per unit cost

[5] McClelland (pp. 373–74) justifies his subsequent duplication of Harper's reliance upon contemporary testimony and attribution of infinite elasticity to Amsterdam demand by distinguishing between his search for an *upper* bound and Harper's search for a *lower* bound.

[6] I am heavily indebted to Robert W. Fogel for elucidating the following in his class in American Economic History at the University of Chicago.

[7] McClelland, pp. 380–81. The exact conditions are given in his fn. 54, p. 381.

imposed on reexports to Amsterdam by the Navigation Acts requiring that enumerated exports pass through England.[8]

Accepting McClelland's contention that "Intensive competition, especially among Scottish factors based in Maryland and Virginia, minimized the opportunity for monopsonistic practices by marketing firms" (p. 378) implies

$$P_{A1} = P_{C1}(1 + f_1 + t)X_1$$
$$P_{A2} = P_{C2}(1 + f_2)X_2$$

whence

$$(2)\qquad \hat{P}_{C0} = \frac{P_{A1}}{P_{A2}} P_{C2} = \frac{P_{C1}(1 + f_1 + t)}{P_{C2}(1 + f_2)} \cdot \frac{X_1}{X_2} \cdot P_{C2}$$

$$(3)\qquad = P_{C1} \cdot \frac{1 + f_1 + t}{1 + f_2} \cdot \frac{X_1}{X_2}$$

in contrast to the "true" value of P_{C0}:

$$(4)\qquad P_{C0} = P_{C1} \cdot \frac{1 + f_1 + t}{1 + f_1}$$

if Amsterdam's demand schedule were of infinite elasticity, where

f_i = the proportional direct freight rate at time i and

X_i = the Amsterdam-colonies exchange rate at time i

Contrary to McClelland's criticisms, Thomas' estimate does not founder on a "... multitude of shifting supply and demand conditions for specific commodities" (McClelland p. 375). Thomas' $\hat{P}_{C0}$ will be greater than the true in-the-absence-of the Navigation Acts price P_{C0}, and thus his estimated burden $P_{C1}\hat{P}_{C0}BA$ greater than the true burden on reexports, if the post-Revolution proportionate direct freight rate f_2 is less than the pre-Revolution rate f_1 and American inflation in the intervening period matched or exceeded Dutch inflation (so that X_1/X_2 is greater than or equal to 1). That is, Thomas' implicit assumption underlying his estimate $\hat{P}_{C0}$ are: 1) the Amsterdam demand schedule was of infinite elasticity; 2) proportional direct freight rates were at least equal in the two periods, $f_1 = f_2$; 3) American inflation at least matched, if not exceeded, Amsterdam inflation in the period; and 4) American supply of exports to Amsterdam was unit elastic.

Thomas' procedure may be generalized to permit non-infinite elasticities of demand and non-unit elasticities of supply. The burden on reexports represented by the area $P_{C1}P_{C0}BA$ in Figure 1c is

$$(5)\qquad L = \int_{P_{C1}}^{P_{C0}} Q_S^C dP_C$$

$$(6)\qquad = \int_{P_{C1}}^{P_{C0}} S P_C^s dP_C$$

$$(7)\qquad = \frac{S}{s+1} P_C^{s+1} \Big|_{P_{C1}}^{P_{C0}}$$

$$(8)\qquad = \frac{Q_1 P_{C1}^{-s}}{s+1} (P_{C0}^{s+1} - P_{C1}^{s+1})$$

$$(9)\qquad = \frac{Q_1 P_{C1}}{s+1} \left(\left[\frac{P_{C0}}{P_{C1}} \right]^{s+1} - 1 \right)$$

If demand is infinitely elastic and supply is unit elastic and freight rates are unchanged, this is just Thomas' estimate of the burden

$$\frac{\hat{P}_{C0}}{P_{C1}} = 1 + \frac{t}{1+f}$$

whence the loss[9]

[8] Thomas assumes that freight costs are proportional to value. McClelland assumes freight charges are absolute per unit (ΔT), although he writes the burden as a percentage of price (σ).

[9] See Thomas, pp. 623–24 and fn. 25, p. 625. In McClelland's symbols and under the above assumptions $L = (x-j)E \cdot \sigma T + (x-j)E \cdot \sigma^2$ where $(x-j)E$ = the value of reexports $= Q_{A1}P_{C1}$ and $\sigma = \Delta P_C/P_{C1} = t/(1+f)$ (McClelland, pp. 380–81). Equation (11)'s equivalence to (12) rests upon the assumed unit elasticity of supply of exports, such that $\Delta Q = Q_2\Delta P/P$.

$$(10)\quad L = \frac{Q_{A1}P_{C1}}{2}\left(\left[1+\frac{t}{1+f}\right]^2 - 1\right)$$

$$(11)\quad = Q_{A1}P_{C1}\left(\frac{t}{1+f}\right) + \frac{Q_{A1}P_{C1}}{2}\left(\frac{t}{1+f}\right)^2$$

$$(12)\quad = Q_{A1}\Delta P_C + \frac{1}{2}\Delta Q_A \Delta P_C$$

More generally, equating $Q_D^C = Q_S^C$ to yield the equilibrium colonial f.o.b. prices P_1^C, P_0^C implies

$$(13)\quad \frac{P_{C0}}{P_{C1}} = \frac{(D/S)^{1/(s_0+d_0)}(1+f)^{-d_0/(s_0+d_0)}}{(D/S)^{1/(s_1+d_1)}(1+f+t)^{-d_1/(s_1+d_1)}} \cdot \frac{X_0^{-d_0/(s_0+d_0)}}{X_1^{-d_1/(s_1+d_1)}}$$

$$(14)\quad = \left(1+\frac{t}{1+f}\right)^{d/(s+d)}$$

if elasticities and the terms of trade are assumed unchanged by the hypothetical disappearance of the Navigation Acts. Then the loss on reexports may be found by substituting (14) into the earlier equation (9).

$$(15)\quad L = \frac{Q_1P_{C1}}{s+1}\left(\left[1+\frac{t}{1+f}\right]^{d(s+1)/(s+d)} - 1\right)$$

A binomial expansion of

$$\left(1+\frac{t}{1+f}\right)^{d(s+1)/(s+d)}$$

to the third term implies

$$(16)\quad L \approx \frac{Q_1P_{C1}}{s+d}\cdot d \cdot\left(\frac{t}{1+f} + \frac{s(d-1)}{2(s+d)}\left(\frac{t}{1+f}\right)^2\right)$$

$$(17)\quad < Q_1P_{C1}\left(\frac{t}{1+f} + \frac{s}{2}\left(\frac{t}{1+f}\right)^2\right) = \operatorname*{Limit}_{d\to\infty} L$$

Hence, the direction of bias in Thomas' substitution of $\hat{P}_{C0}$ for the true P_{C0} in estimating the burden on reexports of colonial goods is clearly upwards if deterioration in the American terms of trade with the Dutch and an elasticity of Dutch demand less than infinity more than offset any appreciation in proportionate direct freight costs over the period. If freight costs declined over the period, as suggested by Douglass North, Thomas' bias is unambiguously upwards on the portion of the loss he considers, contrary to McClelland's allegations (pp. 374–75).

Table 1 calculates the burden as a percentage of reexports for selected elasticity values employing Thomas' implicit estimate of $t/(1+f) = .4$ and assuming no change in the terms of trade or the elasticities stemming from the hypothetical nullification of the entrepôt provisions of the Navigation Acts.[10]

TABLE 1

s \ d	1.0	5.0	∞
.5	27	39	44
1.0	20	38	48
5.0	07	28	80

The burden is calculated as follows:

$$\frac{L}{Q_1P_1} = \frac{d}{s+d}\left[\frac{t}{1+f} + \frac{s(d-1)}{2(s+d)}\left(\frac{t}{1+f}\right)^2\right] \quad \text{for } d < \infty$$

$$= \frac{t}{1+f} + \frac{s}{2}\left(\frac{t}{1+f}\right)^2, \quad \text{as } d \to \infty$$

[10] Thomas estimates a burden of £499,000 on reexports of approximately £950,000 in 1770, which implies (under Thomas' assumptions) that $t/(1+f) \approx .4$. See Thomas, p. 622 and Table 1, p. 626. McClelland estimates $\sigma = \Delta T/P_C = .4$, where ΔT is the absolute unit burden on enumerated reexports to Amsterdam, such that with infinite demand elasticity for such reexports, $\Delta P_C = \Delta T = .4$. If freight rates are *proportional* to P_C (à la Thomas) with the values McClelland indicates, $f = .6$ and $\Delta T = .4$, then $t/(1+f) = .4/1.6 = .25$, for rising freight rates would absorb some of the released unit burden.

Table 2 records the corresponding burden as a percentage of estimated GNP, making use of the following relation:

$$\frac{L}{GNP} = \frac{L}{P_1Q_1}\frac{P_1Q_1}{GNP},$$

where P_1Q_1/GNP is estimated as .025 in 1770. (See Thomas, p. 622.)

TABLE 2

s \ d	1.0	5.0	∞
.5	.7	1.0	1.1
1.0	.5	.9	1.2
5.0	.2	.7	2.0

Increasing Table 2's entries by 14 percent to reflect the colonists' loss on enumerated exports consumed within the Empire (area $P_{C1}P_{C0}MN$ in Figure 1b)—a rough average based on Thomas' assumption that 85.4 percent of enumerated exports to England were reexported and McClelland's assumption that 91 percent were reexported[11]—and then doubling this value in accordance with McClelland's assumption that the burden on imports at most equalled the burden on exports yields illustrative estimates of the burden on the colonists from the trade distortions occasioned by the Navigation Acts as a percentage of GNP assuming various elasticities of Amsterdam supply and demand (Table 3).

TABLE 3

s \ d	1.0	5.0	∞
.5	1.6	2.3	2.5
1.0	1.1	2.1	2.7
5.0	.5	1.6	4.6

Formula: (Table 2)×1.14×2.

Tables 1 through 3 clearly indicates the sensitivity of the total burden to the assumed elasticities and propensities to trade, and support Thomas' and McClelland's contention that the entrepôt requirements of the Navigation Acts couldn't have been too costly for the colonists under a variety of assumptions.[12]

As Harper, Thomas, and McClelland each pointed out, the final calculation on the net costs of the Navigation Acts hasn't been performed.[13] But I hope the above has illustrated the power of a minimal amount of data employed within a consistent and simple model to get around a seemingly insurmountable "missing" datum and yield good answers with clear indications of the source and direction of the biases therein.

REFERENCES

L. A. Harper, "The Effect of the Navigation Acts on the Thirteen Colonies," in R. B. Morris, ed., *The Era of the American Revolution*, New York 1939, 3–9.

———, "Mercantilism and the American

[11] See Thomas (p. 622) and McClelland (p. 397). McClelland figures 55 percent of all exports went to England, of which 50/55ths were reexported. Assuming *all* the colonial exports to England were enumerated implies 91 percent of enumerated exports were reexported. How McClelland's data support these numbers is unclear, for he asserts " . . . 87 percent of American tobacco, by weight, was reexported from Great Britain. Given the dominance of tobacco in total colonial exports to Britain (about 50 percent, 1768–72) and the limited reexport trade in most other goods, 0.5 would seem to be a reasonable upper bound estimate of $(x-j)$" (p. 377). But if tobacco enjoys the highest reexport rate, 87 percent by weight, and other enumerated goods a much lower rate, McClelland's assumption that $(x-j)=.5$ (which implies a reexport rate of 91 percent) appears excessively high. $.24=.55-[1-.87)(.5)+1(.5)](.55)=(x-j)$ is an estimate more consistent with his data; Thomas' comparable estimate is $.36=(.854)(.76)(.55)=(x-j)$. As McClelland says, "The challenge is to measure the percentage of colonial gross national product that was sacrificed because of trade route distortions caused by the Navigation Acts" (p. 374), rather than to generate a very much upper bound.

[12] In the case where the elasticity of demand is infinite and the elasticity of supply is 1, Thomas' corresponding estimate would be 1.2×1.17×2=2.8 percent of GNP. McClelland's would be 3.0×1.10=3.3 percent of GNP.

[13] Although Thomas tentatively estimates an upper bound average net loss of $1.24 per capita in 1770 from all the provisions of the Navigation Acts (Thomas, Table 2, p. 638).

Revolution," *Can. Hist. Rev.*, Mar. 1942, *3*, 1–15.

J. R. T. Hughes, "Discussion of McClelland's Paper," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 382–85.

P. D. McClelland, "The Cost to America of British Imperial Policy," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 370–81.

D. C. North, "Sources of Productivity Change in Ocean Shipping, 1600–1850," *J. Polit. Econ.*, Sept.-Oct. 1968, *76*, 953–67.

R. P. Thomas, "A Quantitative Approach to the Study of the Effects of British Imperial Policy upon Colonial Welfare: Some Preliminary Findings," *J. Econ. Hist.*, Dec. 1965, *25*, 615–38.

On Navigating the Navigation Acts with Peter McClelland: Reply

By PETER D. MCCLELLAND*

Joseph Reid's article would seem to illustrate some of the strengths and weaknesses of what has become known as "the new economic history." First, to the strengths. With the assistance of economic theory, Reid has uncovered a cost of imperial policy neglected by both Robert Thomas and myself: the ". . . redirection of previous production for outside the Empire consumption to English and domestic sale, leading to price declines and attendant loss to the colonists. . . ." The domestic part of that cost would seem impossible to measure, given the absence of information on colonial production and consumption. Its neglect would nevertheless seem defensible, insofar as losses to colonial producers were at least partially offset by gains to colonial consumers. Certainly its inclusion would be unlikely to modify my basic conclusion discounting the importance of income sacrificed because of British interference with colonial trade routes. The British segment of neglected cost is more easily incorporated into the estimation procedure used in my original article. What is required for exports is the value of all enumerated goods as a percentage of total colonial exports. James Shepherd's data indicate that for the period 1768–72, 55 percent of all colonial exports went to Britain.[1] Thomas notes (p. 622) that 76 percent of all American exports to Britain were enumerated goods in 1770. The appropriate revision is therefore to replace $(x-j)=0.5$ (a fraction deliberately inflated to insure an upper bound) by the more accurate $(0.55)\,(0.76)=0.42$.

The second part of Reid's comment is also an admirable exercise in economic theory. The goal is to discover those assumptions required to give validity to the Thomas measurement procedure. To discover the assumptions, however, is not to establish the validity of the procedure. This can only be done by presenting evidence to show the correspondence of those assumptions with the historical reality in question. This Reid never does. Having used the sophisticated tools of the economic theoretician, he fails to apply the simple methodology of the historian. What then are the assumptions, and what is the evidence?

1) American inflation at least matched, if not exceeded, Amsterdam inflation. The evidence is that American inflation was considerably less than Amsterdam inflation.[2]

2) Amsterdam demand schedules were infinitely price elastic. Gray's detailed research on American agriculture suggests that for tobacco—the major colonial export—demand was highly inelastic.[3]

3) American supply schedules for exports to Amsterdam were unit elastic. Direct evidence is extremely scarce, but Gray is prepared to conclude that for colonial tobacco, supply was "essentially inelastic."[4]

4) The post-Revolution proportionate direct freight rate f_2 was less than, or equal to, the pre-Revolution rate f_1. On the availability of freight rate evidence for this period, Douglass North concluded after extensive research that "From the late eighteenth century until the end of the War

* Harvard University.

[1] See McClelland, Appendix C.

[2] A comparison of price indexes for the period 1785–94 with those of 1765–74, for example, suggests that on the average Amsterdam prices were 28 to 41 percent higher in the post-Revolutionary period while those of America were 18 to 22 percent higher. (See N. W. Posthumus, Vol. I, p. ci; Anne Bezanson et al. pp. 388–90; U.S. Bureau of the Census, p. 116.)

[3] (L. C. Gray, Vol. 1, pp. 257–58; Vol. II, p. 762).

[4] (Gray, Vol. I, p. 276). Thomas actually assumes an elasticity of supply of 0.5 for rice and 1.0 for all other commodities examined. Only in the case of tobacco does he give supporting evidence (Thomas, p. 624, n. 24.), and even that is unacceptable unless one is prepared to assume no shifts whatsoever in the colonial supply function between 1763–72 and 1790–93.

of 1812, the freight rate data are so influenced by recurrent wars that it has not been possible to construct an American series" (p. 954). In no case, therefore, does the historical reality appear to correspond with the Thomas assumptions. As for the critical issues of freight rates and relative price inflations, available evidence is either nonexistent or directly contradicts the assumption required.

The final section of Reid's paper attempts to generalize the Thomas procedure under different assumptions concerning supply and demand elasticities. The cost estimates so derived are then expressed as a percentage of colonial Gross National Product, following Thomas' calculation of re-exports equalling about 2-1/2 percent of *GNP* in 1770. Thomas' sources for *GNP* estimates are merely the rough impressions of other economic historians:

> Albert Fishlow stated at last year's meetings that he believed that the average per capita income in the 1780's "could not have been much less than $100." George Rogers Taylor, in his presidential address, hazarded a guess that per capita income did not grow very rapidly, if at all, between 1775 and 1840. [p. 638]

From this Thomas ventures the assumption that "average per capita income hovered about $100 between 1763 and 1772" (p. 638). The tenuous nature of this *GNP* estimate and the dubious validity of assumptions underlying the Thomas procedure would seem in concert to discount the value of Reid's final cost calculations.

One misunderstanding remains to be clarified. My own estimate of the economic burden occasioned by the Navigation Acts is the product of three ratios. One of these—an upper-bound estimate of the percentage of colonial exports and imports artificially routed through Britain—was based on Shepherd's reworking of original colonial trade data for the period 1768–72. A second—the ratio of exports plus imports to colonial *GNP*—was based upon a similar ratio for the 1834–43 period, plus associated arguments suggesting why a slightly inflated version (0.15 instead of 0.129) was a reasonable upper-bound estimate for America of pre-Revolutionary days. It is the third which Reid appears to misinterpret. This attempts to estimate for tobacco the incremental costs of routing goods through Britain as a percentage of colonial f.o.b. price. Three independent estimates of these incremental costs were discovered for the years 1729, 1730, and 1736; all in close agreement on approximate magnitudes, at least two from sources unlikely to understate the expenses involved. All three sources gave the cost of transatlantic freight, F, but not the additional cost of moving tobacco from England to northern Europe, ΔF. It was to solve this data shortage that evidence on the *ratio* of $\Delta F/F$ was assembled from 1712–18, the 1730's and the 1760's. That evidence suggested that for the three years in question, 1729, 1730, and 1736, an assumed ratio of $\Delta F/F = 0.25$ was a reasonable upper-bound approximation. This ratio was applied to reported transatlantic freights to derive incremental freight costs. The sum of all incremental costs was then divided by colonial f.o.b. price, giving a ratio of 0.4. The remainder of my Appendix B attempts to demonstrate (a) why this ratio of 0.4 for tobacco shipments of the early 1730's could also be used for tobacco shipments of the early 1770's, and (b) why the estimated ratio for tobacco could also be used as an acceptable upper-bound proxy for all colonial exports and imports. The end result is far from perfect, relying as it does on fragmentary evidence from the eighteenth century. The key methodology used to bind that evidence together, however—the extrapolation of ratios from one period to another—is surely miscast as a "following of Harper's previously criticized footsteps."

Alfred Marshall's *Principles of Economics* begins with the motto "Natura non facit saltum." Perhaps the new economic historian should also beware of proceeding by leaps. Armed with the tools of Marshall and others, he can explore where others have not been. But novel tools do not alone assure success. At his peril he ignores the *caveats* of more conventional historians: how good are

the sources? how adequate is the evidence? The theoretical constructs of this newer breed of analyst promise much, but only insofar as assumptions implicit in his theorizing can be supported by historical evidence. And data—that weft and warp of almost all his labors—are no better than the sources from which they come.

REFERENCES

A. Bezanson et al., *Wholesale Prices in Philadelphia,* Philadelphia 1936.

L. C. Gray, *History of Agriculture in the Southern United States to 1860,* Washington 1933.

P. D. McClelland, "The Cost to America of British Imperial Policy," *Amer. Econ. Rev. Proc.,* May 1969, *59,* 370–81.

D. C. North, "Sources of Productivity Change in Ocean Shipping, 1600–1850," *J. Polit. Econ.,* Sept./Oct. 1968, *76,* 953–70.

N. W. Posthumus, *Inquiry into the History of Prices in Holland,* Leiden 1946.

R. P. Thomas, "A Quantitative Approach to the Study of the Effects of British Imperial Policy Upon Colonial Welfare," *J. Econ. Hist.,* Dec. 1965, *25,* 615–38.

U.S. Bureau of the Census, *Historical Statistics of the United States,* Washington 1960.

Tariffs, Intermediate Goods, and Domestic Protection: Comment

By BELA BALASSA*

We are indebted to Roy Ruffin for having provided us, in a recent issue of this *Review*, with a simple, yet elegant, general equilibrium model incorporating intermediate goods, which is suitable for an analysis of protection. However, as I will show below, Ruffin has erroneously discarded the current measure of effective tariff that has general applicability, replacing it by a measure whose validity is limited to a special case even under Ruffin's proposed definitions. At the same time, these definitions are based on restrictive conditions which are both unnecessary for the model and at variance with effective protection theory. Yet Ruffin's principal conclusions crucially depend on the definitions applied and are invalidated if a more general formulation is used. These conclusions are:

1) The current measure of effective tariff, called "the Barber-Johnson-Balassa formula . . . is correct only under the special assumption that exports do not use the imported intermediate goods" (p. 268).

2) "As long as tariffs on intermediate goods are lower than on final goods, the existing estimates should be overestimates of the correct effective tariff rates" (p. 268).

3) Whatever the formula applied, "there is no unique measure of protection in any one industry. Each effective tariff rate depends on the base industry to which the industry in question is compared and two different bases may lead to entirely different conclusions" (p. 268).

4) "When imported intermediate goods *are* produced at home, the foundations of the effective tariff concept seem to disappear completely. If the strict unbiasedness and efficiency properties were the only reasons for using effective tariffs, the whole concept would have to be buried" (pp. 268–69).

* Professor of political economy at the Johns Hopkins University. This note was prepared in the framework of the Project on the Structure of Protection in Developing Countries, financed by the International Bank for Reconstruction and Development and the Inter-American Development Bank, and directed by the author in his capacity as consultant to the economics department of the *IBRD*.

I

In the following, I will disregard the consumption side of the model, and will consider only the conditions for productive efficiency. Accordingly, I have omitted A.2., the assumption pertaining to the utility function. The remaining assumptions are:

- A.1. There are three goods: X_1 and X_2 are final goods produced domestically and X_3 is an imported intermediate good.
- A.3. Each unit of X_2 requires a_2 units of X_3, i.e., $X_3=a_2X_2$.
- A.4. The gross transformation curve, which defines production possibilities on the assumption the requisite amount of X_3 has been obtained, is concave to the origin and is given by $T(X_1, X_2)=0$.
- A.5. The prices P_1, P_2, and P_3 are fixed on world markets. Later, we shall replace assumption A.3 by:
- A.3′. Each unit of $X_i(i=1, 2)$ requires a_i units of X_3.

Under assumptions A.1. through A.5., the contribution of a marginal unit of X_1 to national income (value-added per unit of output or the net price of X_1) will equal the price of the product ($P_1'=P_1$) while for X_2 this is defined as[1]

$$(1') \qquad P_2' = P_2 - a_2P_3$$

[1] I have used the symbols employed by Ruffin except that I have denoted the effective rate of tariff on the ith good by z_i. For easy identification, all my equations have a primed number.

The condition for productive efficiency can be derived by equating the relative contributions to income of the two domestically produced goods with the marginal rate of transformation between them:

$$(2') \qquad T_1/T_2 = P_1'/P_2' = P_1/P_2'$$

Ruffin next advances the following definitions:

D.1. Let P_i' be the domestic value-added or net price of the jth good; and let the kth final good be free of subsidies and tariffs. Then a measure, t_i', is said to be the effective tariff rate (relative to k) on industry i if t_i' is positive, zero, or negative, if and and only if, the marginal rate of transformation, T_i/T_k, exceeds, equals, or falls short of P_i'/P_k'.

D.2. If when the effective tariff rate on the imported final good is zero and the tariff on the final good is nonzero, the production point of the economy is the same as the production point under free trade, then the effective tariff is said to be *unbiased.*

D.3. If when the effective tariff on the imported final good is zero and the tariff on the final good is nonzero, a second-best optimum is achieved, then the effective tariff is said to be *efficient.*

For unbiasedness and efficiency, the definitions put forward by Ruffin thus require (a) the tariff on a kth final good (the export product in his example) to be zero; (b) the tariff on the imported final good to be nonzero; and (c) the effective tariff on the imported final good to be zero. But these restrictive conditions are not necessary for the model and are at variance with effective protection theory. Rather than imposing restrictions on the absolute values taken by nominal tariffs (subsidies) and effective rates, the theory has been formulated in terms of ratios of effective rates.

As W. M. Corden notes, "The scale [of effective tariffs] summarizes the total protective-rate structure. Assuming normal nonzero substitution elasticities, it tells us the direction in which this structure causes resources to be pulled as between activities producing traded goods" (p. 224). The corollary of this proposition is that such a resource flow—taking us away from the free trade production point—will not occur if effective rates on all activities are equal. In other words, the equality of effective rates of tariffs is a condition for productive efficiency.

Relaxing the restrictive conditions imposed by Ruffin, we obtain a general formulation of the definitions:

D.1′ Let P_j' be the domestic value-added or net price of the jth good. Then a measure, z_i, is said to be the effective tariff rate (relative to the kth final good) on industry i if z_i is greater than, equal to, or less than z_k, if and only if, the marginal rate of transformation, T_i/T_k, exceeds, equals, or falls short of P_i'/P_k'.

D.2.′ If under protection, the production point of the economy is the same as the production point under free trade, then the effective tariff is said to be *unbiased.*

D.3.′ If under protection, a second-best optimum is achieved, then the effective tariff is said to be *efficient.*

II

In Case 1, $t_1 = 0$, $t_2 > 0$, and t_3 is any number. Since the export good (X_1) is not protected and it does not use the intermediate good as an input, its effective tariff will be zero. In turn, the effective tariff on the protected final good, X_2, under the Barber-Johnson-Balassa definition will be

$$z_2 = \frac{P_2(1 + t_2) - a_2P_3(1 + t_3)}{P_2 - a_2P_3} - 1$$

$$(3') \qquad = \frac{P_2t_2 - a_2P_3t_3}{P_2 - a_2P_3}$$

According to our definitions, with z_1 equal to zero, productive efficiency requires that z_2, too, should be zero. Equation (5′) shows the tariff on the intermediate good, t_3, that is compatible with the condition. This equa-

tion is identical to Ruffin's equation (5), derived from the conditions for a second-best optimum. The Barber-Johnson-Balassa measure of effective tariff is thus unbiased and efficient.

$$(4') \qquad z_1 = z_2 = \frac{P_2 t_2 - a_2 P_3 t_3}{P_2 - a_2 P_3} = 0$$

$$(5') \qquad t_3 = \frac{t_2 P_2}{a_2 P_3}$$

In Case 2, it is assumed that $t_2 = 0$, $t_1 > 0$, and X_2, the commodity using the imported intermediate good as an input, rather than X_1, is exported. The effective rates of tariff on X_1 and X_2 are, then, expressed by (6′) and (7′) while (8′) shows the condition for the equality of effective tariff on these goods.

$$(6') \qquad z_1 = \frac{P_1(1 + t_1)}{P_1} - 1 = t_1$$

$$(7') \qquad z_2 = \frac{P_2 - a_2 P_3(1 + t_3)}{P_2 - a_2 P_3} - 1$$

$$= -\frac{a_2 P_3 t_3}{P_2'}$$

$$(8') \qquad z_1 = t_1 = -\frac{a_2 P_3 t_3}{P_2'}$$

Solving for t_3, the tariff on the intermediate good in equation (9′), we find that this equals the second-best tariff derived from optimum conditions in Ruffin's equation (10). Thus the Barber-Johnson-Balassa measure of effective tariff is again unbiased and efficient.

$$(9') \qquad t_3 = -\frac{t_1 P_2'}{a_2 P_3}$$

While Ruffin expressed surprise that in this case the second-best tariff on the intermediate good is negative, the result can be easily explained. The protection of X_1 introduces discrimination against the export good, X_2, and in the absence of a direct subsidy on the exports of X_2, this discrimination can be offset only by subsidizing the imports of X_3 that are used in its production.

In Case 3, $t_1 = 0$, $t_2 > 0$, and the imported intermediate good is used in both the production of X_1, the export commodity, and in that of X_2, the protected domestically produced final good (i.e., assumption A.3 is replaced by A.3.′). The effective rates of tariffs on X_1 and X_2 will accordingly be:

$$(10') \quad z_1 = \frac{P_1 - a_1 P_3(1 + t_3)}{P_1 - a_1 P_3} - 1$$

$$(11') \quad z_2 = \frac{P_2(1 + t_2) - a_2 P_3(1 + t_3)}{P_2 - a_2 P_3} - 1$$

Replacing $P_1 - a_1 P_3$ by P_1' and $P_2 - a_2 P_3$ by P_2', equating the effective tariffs on X_1 and X_2, and rearranging terms, we obtain the condition prescribed by Ruffin for a second-best optimum:

$$(12') \quad \frac{P_1'}{P_2'} = \frac{P_1 - a_1 P_3(1 + t_3)}{P_2(1 + t_2) - a_2 P_3(1 + t_3)}$$

Further, expressing t_3 from (12′) we get (13′) which is identical to Ruffin's equation (13). The Barber-Johnson-Balassa measure of the effective tariff is thus unbiased and efficient as in the previous two cases. This result is in conflict with Ruffin's assertion that "... the effective tariff as currently measured, will ... miss the mark completely on both protection and welfare grounds; moreover, it will not satisfy our reasonable definition of the effective tariff rate" (p. 266).

$$(13') \qquad t_3 = -\frac{t_2 P_2 P_1'}{P_3(a_1 P_2 - a_2 P_1)}$$

The difference in the conclusions is explained by the fact that Ruffin imposed the unnecessarily restrictive condition for unbiasedness and efficiency that the effective tariff rate on the imported final good be zero. In Cases 1 and 2, Ruffin's restrictive condition did not cause trouble since the effective tariff on the export commodity was zero. In Case 3, however, the export commodity uses the imported intermediate good as an input. A positive tariff on the imported intermediate good will thus give negative effective protection to the export

commodity, and productive efficiency will require negative effective protection on the domestically produced domestic commodity also. This result may, or may not, require negative tariffs (import subsidies) on X_2, depending on the relative magnitudes of the input coefficients and the prices of the domestically produced commodities.

What then of the measure of effective tariff proposed by Ruffin? Apparently, this has been derived by rearranging the terms of equation (13) on one side and arbitrarily calling it "the effective tariff on X_2". But the result could have been given any other name and the application of the formula does not ensure the equality of effective tariffs on X_1 and X_2, which is a condition of productive efficiency according to the theory of effective protection. Moreover, even if we were to accept the definitions suggested by Ruffin, and require the effective tariff on the imported good to be zero, his measure of effective protection will not give the correct answer in Cases 1 and 2 above.[2]

In Case 4, the intermediate good is protected as well as produced domestically, and it is used in the production of X_2, the domestically produced and protected final good. Under definitions D.1′ through D.3′, unbiasedness and efficiency require that effective tariffs on all domestically produced commodities be equal. Thus, we should have $z_1=z_2$ and $z_1=z_3$, when z_1 and z_2 are as in Case 1 while z_3 equals t_3 since the intermediate good does not use other intermediate goods as inputs. Accordingly,

$$z_1=t_1=z_2=\frac{P_2(1+t_2)-a_2P_3(1+t_3)}{P_2-a_2P_3}-1 \tag{14'}$$
$$=\frac{P_2t_2-a_2P_3t_3}{P_2-a_2P_3}$$

$$z_1=t_1=z_3=t_3 \tag{15'}$$

Utilizing equation (15′) and expressing t_3 from equation (14′) we get $t_3=t_2$. It follows that, if all commodities used domestically are produced at home, the equality of effective tariff rates presupposes the equality of nominal tariffs (subsidies) on all commodities. Identical results are obtained from the appropriate conditions for the second-best optimum:

$$\frac{P_1}{P_2'}=\frac{P_1(1+t_1)}{P_2(1+t_2)-a_2P_3(1+t_3)} \tag{16'}$$

$$\frac{P_1}{P_3}=\frac{P_1(1+t_1)}{P_3(1+t_3)} \tag{17'}$$

From equation (17′) we find that the second-best optimum requires the equality of t_1 and t_3. Substituting t_1 for t_3 in equation (16′), the same requirement is obtained in regard to t_2 and t_3. Thus, we have shown that the Barber-Johnson-Balassa measure is both unbiased and efficient, while according to Ruffin it is biased as well as inefficient. Ruffin's result has been due to the restrictive assumption that exports be neither taxed nor subsidized. Under the general formulation I have suggested, Ruffin's conclusions are rejected and the desirable properties of the effective protective measure reestablished.

The conclusions are not affected if we assume that the intermediate good is also used in the production of the export commodity, X_1 (Case 5). In this case, we utilize equations (18′) and (19′) to express the equality of z_1 and z_2 and that of z_1 and z'. From (18′), it is apparent that

$$\frac{P_1(1+t_1)-a_1P_3(1+t_3)}{P_1-a_1P_3}-1 \tag{18'}$$
$$=\frac{P_2(1+t_2)-a_2P_3(1+t_3)}{P_2-a_2P_3}-1$$

$$z_3=t_3=\frac{P_1(1+t_1)-a_1P_3(1+t_3)}{P_1-a_1P_3}-1 \tag{19'}$$
$$=\frac{P_1t_1-a_1P_3t_3}{P_1-a_1P_3}$$

[2] Comparing equations (17) and (18), we find that Ruffin's measure will not show zero effective tariffs if such a result is obtained by the use of the Barber-Johnson-Balassa measure unless $t_3=t_2$. Yet, in Cases 1 and 2, unbiasedness and efficiency require *both* that the effective tariff be zero and that nominal tariffs on the two goods be different. At the same time, a zero effective tariff has been obtained in both cases by the Barber-Johnson-Balassa measure.

The equality of z_1 and z_3 requires the equality of t_1 and t_3. Substituting t_3 for t_1 in (18′) and rearranging terms we obtain (20′). Again, the equality of effective rates will require the equality of tariffs on X_2 and X_3.

$$(20') \quad \frac{P_1 t_3 - a_1 P_3 t_3}{P_1 - a_1 P_3} = \frac{P_2 t_2 - a_2 P_3 t_3}{P_2 - a_2 P_3}$$

The conditions for second-best optimum will give the same results.

$$(21') \quad \frac{P_1'}{P_2'} = \frac{P_1(1 + t_1) - a_1 P_3(1 + t_3)}{P_2(1 + t_2) - a_2 P_3(1 + t_3)}$$

$$(22') \quad \frac{P_1'}{P_3} = \frac{P_1(1 + t_1) - a_1 P_3(1 + t_3)}{P_3(1 + t_3)}$$

The equality of t_1 and t_3 is derived from (22′) while, substituting t_3 for t_1 into (21′), the same requirement is obtained with regard to t_2 and t_3. Thus, the Barber-Johnson-Balassa measure is shown to be unbiased and efficient also in the more general case where both final goods use the intermediate good as an input and the latter is produced domestically. It is of further interest that we have reestablished the traditional conclusion, obtained in models incorporating final goods only that productive efficiency requires the equality of all tariffs and subsidies for the case when intermediate goods are introduced.

Under definition D_1 we also find that there is a unique ranking of commodities by effective rates. All pairwise comparisons are consistent with this ranking and hence it does not matter which is the "base industry" used in such comparisons. This conclusion again differs from Ruffin's by reason of the fact that he has imposed the unnecessarily restrictive condition that the base industry have zero nominal tariff, although no such condition is found in effective protection theory.

In conclusion, I have shown the following propositions to be correct:

1) The validity of the Barber-Johnson-Balassa measure of effective tariff does not depend on the special assumption that exports do not use intermediate goods;
2) The measure is unbiased and efficient in a general equilibrium model incorporating intermediate goods;
3) The Barber-Johnson-Balassa measure provides a unique ranking of industries according to their effective rates of protection;
4) The unbiasedness and efficiency properties of the effective rates of tariffs are not affected if imported intermediate goods are produced domestically.

These conclusions reaffirm the validity of effective protection theory and indicate the unbiasedness and efficiency properties of the Barber-Johnson-Balassa measure of effective rates in a general equilibrium framework. It is a different question that the model utilized does not admit substitution among inputs and thus avoids the controversial topic of bias in measurement due to input substitution. But this is another story.

REFERENCES

W. M. Corden, "The Structure of a Tariff System and the Effective Protective Rate," *J. Polit. Econ.*, June 1966, *74*, 221–37.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Production," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

Tariffs, Intermediate Goods, and Domestic Protection: Reply

By Roy J. Ruffin*

I cannot agree with Bela Balassa that *my* conclusions change if different definitions are used. The conclusions following from new definitions are *new* conclusions, not mine. The issue is whether the conclusions are interesting and correct.

A careful reading of my paper will indicate that in all cases I look at the second-best optimum for a *given* consumption distortion. For the case where the intermediate good was not produced at home, it was shown that one can find a nominal tariff on the intermediate good such that a second-best optimum can be achieved for a specified consumption distortion. For the case where the intermediate good was produced at home, it was impossible to find a tariff on the intermediate good that would lead to a second-best optimum. Balassa, on the other hand, looks at the second-best solutions for any consumption distortion, including the absence of a consumption distortion. It is not surprising, therefore, that he reaches new conclusions. But the "new" conclusions reinforce, rather than contradict, my results.

Assume a world in which ad valorem import tariffs and export subsidies (collectively referred to as tariffs) are the only impediments to the competitive adjustment of markets, all goods are traded, and the prices of all goods are fixed on world markets. If all intermediate goods have fixed coefficients of production, the production functions involving the primary factors and the supplies of those factors can be used to construct the gross transformation function $T(X_1, \ldots, X_m)=0$, where X_i = the amount of the ith good produced at home. For a given consumption distortion, a second-best optimum requires that production be at the free trade production point:

$$(1)\quad T_i/T_k = P_i/P_k' \quad i \neq k;\ i = 1, \ldots, m,$$

where P_j' is the price of good j net of the cost of intermediate goods. If z_i and z_k are the Barber-Corden-Johnson-Balassa effective tariff rates on goods i and k, which measure the percentage changes in the domestic value-added of the two goods, then under competition it will happen that

$$(2)\quad T_i/T_k = P_i'(1+z_i)/P_k'(1+z_k), \quad i \neq k;\ i = 1, \ldots, m$$

Thus, *if* there is a second-best optimum for a given consumption distortion, it must be that $z_i = z_k$ for all i. But one may not be able to set $z_i = z_k$ for all i without changing the original consumption distortion. This is actually demonstrated by Balassa for the case of two produced final goods and one produced intermediate good. In that case, Balassa shows that $z_1 = z_2 = z_3$ implies a *uniform* tariff rate or export subsidy, which means there is *no* consumption distortion (indeed, it is free trade). In other words, for any consumption distortion a second-best optimum cannot be achieved by merely manipulating effective tariff rates.

Balassa's proposition that $z_i = z_k$ for all i leads to a second-best optimum must not be misinterpreted. If it does not lead to a first-best optimum (as above), it may actually be *worse* than some tariff-distorted state that does not satisfy the conditions for a second-best solution. To have a second-best optimum means that welfare is greater than it would be under a tariff-distorted production point with the *same* consumption distortion. If a second-best optimum is compared to a tariff-distorted production point with a different consumption distortion, no general conclusion is possible.

This is illustrated in Figure 1 for the case of two produced final goods, X_1 and X_2, and one unproduced intermediate good, X_3, used

* Associate professor of economics at The University of Iowa.

in the production of X_2. TT' is the gross transformation curve (defined by ignoring requirements of intermediate good), RR' gives the combinations of X_1 and X_2 consistent with a given level of national income (the weights are domestic value-added for each good), and Rr gives the consumption possibilities corresponding to the maximum level of national income. The points E and e give the first-best optimum. The second-best optima are given by all the points along Rr, excluding e. A representative second-best solution is given by the point e' and the indifference curve u_o. If any tariff-distorted equilibrium has a consumption point in the interior of the shaded region bounded by Rr and u_o, that equilibrium will be better than the representative second-best solution. Eliminating protection does not necessarily make a country better off.

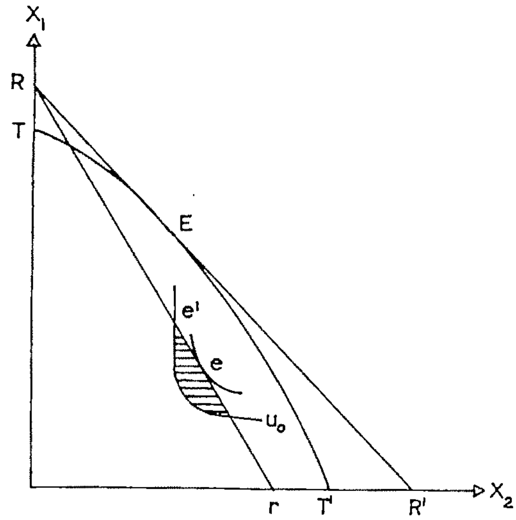


FIGURE 1

Balassa is also confused over the role of my definition of the effective tariff on good i relative to good k. He asserts, incorrectly, that under his (slightly generalized) definition this turns out to be the z_i measure. The relative measure is zero if, and only if, $T_i/T_k = P_i'/P_k'$. This implies, from (2), that the relative measure will be zero if, and only if, $z_i = z_k$. It is clear, therefore, that the relative effective tariff rate is not the absolute effective tariff rate, z_i. Balassa's allegation that my definitions of unbiasedness and efficiency require z_i to be zero is incorrect; I require the relative measure to be zero.[1]

The purpose of the relative measure is to focus on the fact that it is relative prices, not absolute prices, that direct the allocation of resources. This is clear from equation (2). Just as relative prices do not displace absolute prices as interesting concepts, the relative measure does not displace the absolute measure. Indeed, I think that the term "effective tariff rate" should mean the absolute measure (though it was not used this way in my article), and if one wants to discuss the relative measure, it should be called the relative effective tariff.

Unfortunately, in my paper I conjectured that the relative effective tariff rate need not give the same ranking of industries as the effective tariff rate. This turns out to be false for the relative measure proposed in my paper, but not for all relative measures. But this is not a critical issue.

The important question is whether ranking industries by effective rates imparts knowledge about degrees of protection. This problem was not systematically discussed in my paper, but it was hinted that production-substitution relationships would have to be explored. If, of course, we define protection by the effective tariff rate, as many people do, that begs rather than answers the question. The purpose of the ranking is to make inferences about the resource-allocation effects of an entire tariff structure. It has been presumed that resources will move from high to low effective rate activities, but this has not been proved.

Why does this presumption need proof? Suppose there are m domestically produced goods. Let these goods be labeled so that $z_1 > z_2 \ldots > z_m$. Since only relative values matter, we can impose the restriction that $z_1 = 0$. Thus, for any ranking of effective tariffs, we can think of the effects on resource allocation as being identical to holding the

[1] W. M. Corden (1969) also introduces the notion of the relative effective rate.

(net) price of the first good fixed and reducing all the other (net) prices by varying percentages. It seems clear that this will discourage the production of at least one of the other goods, but whether it will encourage the production of the first good is far from obvious. From the composite commodity theorem we know that if all other prices were reduced by the *same* percentages, the supply of the first good would have to be stimulated. But if there are reduced by *varying* percentages, we cannot appeal to this famous Hicksian theorem. To assert only that the supply of the first good is stimulated (which is weaker than the assertion that resources move from high to low effective rate activities) is to assert a theorem of a higher rank than the composite commodity theorem.[2] If such a theorem were true under general conditions, i.e., concavity of the production-possibility frontier, it would be one of the most remarkable theorems in value theory. Unfortunately, it is probably false, as could be shown by a counterexample, which I have not been able to produce.

The important contribution of Balassa's paper, lest it be missed in the haze of controversy, is that he forces upon us the realization that for the model proposed in my article a second-best optimum requires equality of all effective tariff rates. This was implicit in my own paper, as Balassa himself shows. But Balassa states the theorem awkwardly and ends up misinterpreting its significance. To prevent this mistake, I suggest the theorem be stated as follows: if there is a second-best optimum for a given consumption distortion, to reach that optimum all effective tariff rates must be equal and the original consumption distortion cannot be disturbed. This theorem clarifies the nature of my own contribution and provides the link between my work and previous efforts. For example, my definitions of unbiasedness and efficiency were not formulated in the light of this theorem and, hence, their meaning is unclear. They were formulated, however, to discuss a state of affairs: the existence or nonexistence of a true second-best optimum. I would now redefine these terms as follows: if under *all* consumption distortions (unequal nominal tariff rates on consumer goods), it is possible to reach the free trade production point, the effective tariff is said to be unbiased and efficient.[3] It is said to be unbiased since production is not distorted; it is said to be efficient since a second-best optimum is achieved. Since the ranking of effective tariff rates yields little information about resource allocation, the unbiasedness and efficiency property would be a nice theoretical foundation. But, in the model considered in my paper, if the intermediate good is domestically produced, the effective tariff is biased and inefficient, and free trade production cannot be attained without going all the way to free trade. In this case there are *no* consumption distortions consistent with the free trade production point. In more general models, it is possible for some but not all consumption distortions to be consistent with the free trade production point.

The above conclusions also hold for a given production distortion, increasing the practical importance of the results. For example, in Figure 1 any point to the southeast of E can be reached by an import tax on X_2 or an import subsidy on X_3, the intermediate good. The latter will be better since it does not involve a consumption distortion.

The failure of effective tariffs to satisfy the unbiasedness and efficiency properties (as defined above) does not, logically, con-

[2] Corden (1966, p. 224) asserts two distinct propositions about resource allocation: (i) resources move from low to high effective rate activities; and (ii) resources move from the lowest to the highest effective rate activity, with no prediction about other activities. The strong proposition (i) is false if the weak proposition (ii) is false. In an earlier draft of this paper, I argued against the strong proposition. I am indebted to my colleague, Professor Darwin Wassink, for inducing me to question the weak proposition.

[3] Balassa's definitions of unbiasedness and efficiency are, as stated, meaningless. Consider his definition of efficiency: "if under protection, a second-best optimum is achieved, then the effective tariff is said to be *efficient*." But Balassa fails to define "if under protection" and finally takes it to include the case of uniform tariffs on *all* goods (free trade), which seems to me to destroy the usefulness of the definition.

stitute rejection of the concept.[4] Indeed, the failure of the ranking of effective tariffs alone to indicate the production distortions of tariffs does not mean the concept ought to be rejected. Effective tariffs influence the pattern of resource allocation and must be considered in any theory of protection; but to announce a definition of effective tariffs is not to announce a theory of protection. Much more work remains to be done.

[4] But, as stated in my article on page 269, if these were "the only reasons for using effective tariffs, the whole concept would have to be buried." The purpose of this statement was to stimulate further research intô effective tariffs, not to condemn the concept.

REFERENCES

W. M. Corden, "The Structure of a Tariff System and the Effective Tariff Rate," *J. Polit. Econ.*, June 1966, *74*, 221–37.

———, "Effective Protective Rates in General Equilibrium: A Geometric Note," *Oxford Econ. Pap.*, July 1969, *21*, 135–41.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Protection," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

Tariffs, Intermediate Goods, and Domestic Protection: Further Comment

By Bela Balassa

In my comment I have shown that the definitions advanced in Ruffin's paper are based on an arbitrary and overly restrictive concept of the effective tariff. Replacing this by the Barber-Johnson-Balassa measure Ruffin rejected, the conclusions he arrived at are invalidated and the new conclusions reaffirm the validity of effective protection theory.

In his reply, Ruffin asserts that "the 'new' conclusions reinforce, rather than contradict, [his] results" (p. 964). But one looks in vain for evidence on this point. Indeed, Ruffin now adopts the Barber-Johnson-Balassa measure of effective protection, and accepts the conclusion that in the models considered "a second-best optimum requires equality of all effective tariff rates", adding the curious postscript that "this was implicit in my own paper, as Balassa himself shows" (p. 966).

Elsewhere Ruffin suggests that his original measure was a "relative effective rate" and states: "Unfortunately, in my paper I conjectured that the relative effective rate need not give the same ranking of industries as the [absolute] effective rate. This turns out to be false for the relative measure proposed in my paper, but not for all relative measures" (p. 965). No proof is provided for this proposition, but none could have been since the expression "relative effective rates" refers to the ratio of absolute effective rates (see W. M. Corden, p. 139), so that the two must give the same ranking. At any rate, it is apparent from equation (17) in the original article that Ruffin's effective tariff measure was defined in absolute and not in relative terms as he alleges. A cursory look at definitions D.2–D.3 and the manipulation of equation (13) will further reveal that Ruffin's definitions of unbiasedness and efficiency require his effective tariff on commodity i to be zero.

It is also incorrect to assert that I would have looked "at the second-best solutions for any consumption distortion" (p. 964). In fact, I followed Ruffin in starting out with a given consumption distortion and then indicated that equalizing effective tariffs will bring forth welfare improvements in all four cases considered. In cases when the intermediate good is not produced domestically, productive efficiency is reached while maintaining the consumption distortion. And, in the fourth case when all commodities are produced domestically, I "have reestablished the traditional conclusion . . . that productive efficiency requires the equality of all tariffs and subsidies" (p. 963).

Ruffin suggests however that "eliminating protection [more precisely, equalizing effective rates] does not necessarily make a country better off" (p. 965). This is not so in the cases considered in Ruffin's original article and his reply offers no proof; the diagram utilized refers to the case when a second-best optimum is compared to a tariff-distorted production point with a different consumption distortion. That productive efficiency with a given consumption distortion may be inferior to a situation when both production and consumption are distorted has been known since Richard Lipsey and Kelvin Lancaster and it does not require proof.

Finally, I turn to the new definition proposed by Ruffin, according to which "if under *all* consumption distortions (unequal nominal rates on consumer goods), it is possible to reach the free trade production point, the effective tariff is said to be unbiased and efficient" (p. 966). This definition is irrelevant for judging the usefulness of the

effective tariff concept because it imposes an unnecessary constraint on it. If all goods are produced domestically, the suggested condition cannot be fulfilled since equalizing effective rates will lead to equal nominal rates. Such a result, however, in no way diminishes the usefulness of the effective tariff concept.

REFERENCES

W. M. Corden, "Effective Protective Rates in General Equilibrium: A Geometric Note," *Oxford Econ. Pap.*, July 1969, *21*, 135–41.

R. G. Lipsey and K. J. Lancaster, "The General Theory of Second Best," *Rev. Econ. Stud.*, 1956 No. 1, *24*, 11–32.

The Cost of Ending the Draft: Comment

By BENJAMIN P. KLOTZ*

Anthony Fisher's recent article in this *Review* is notable as the first attempt to derive a supply curve of military volunteers from an underlying economic theory. Assuming a lognormal distribution of civilian opportunities, he argues for a military supply curve of decreasing elasticity.[1] Allowing for draft pressure, unemployment, and seasonality he derives and fits a time-series supply equation using 1957–65 quarterly data on Department of Defense (*DOD*) enlistments:

$$
\begin{aligned}
E/P = {} & b_0 + b_1 \ln (W_c/W_m) \\
& + b_2 \ln (1 - U/P) \\
& + b_3 \ln (1 - A/P) \\
& + \text{seasonal dummies},
\end{aligned}
\tag{1}
$$

where P = male civilians aged 17 to 20 years; E = enlistments in mental categories I, II, and III; E/P = enlistment rate; W_c/W_m = median civilian wages to average military wages; U/P = youth unemployment rate; A/P = military accession rate.

Fisher shows that in (1) the supply elasticity, ES, equals $b_1/(E/P)$. His estimated elasticities of .46 with a draft and .74 without are consistent with previous cross-section estimates. However, setting draft pressure (inductions) equal to zero in his equation implies that only 24 percent of enlistments are draft-induced, a finding that he correctly questions. He then uses equation (1) to project both the enlistment rate in a draft-free world and the wage increase necessary to raise this rate sufficiently to stock a 2.65 million-man force in 1965. His incremental cost estimates to establish voluntarism in 1965 are between \$4.8 and \$6.9 billion, depending on the youth unemployment rate and the increase in comparable civilian wages.

Proceeding from the specific to the general, several questions may be raised about the methods used:

1. Are the key supply parameters of equation (1) reliably estimated?
2. What is the cost of establishing voluntarism in 1970 rather than in 1965?
3. Are Fisher's results consistent with his theory of supply?

We will see that the answers to questions 1 and 3 must be answered in the negative and consequently the budget cost of voluntarism has been overestimated. Costs appear to be only \$2 billion, far less than Fisher's estimates.

I. *Estimating the Military Supply Curve*

Although Fisher shows that his supply elasticity of .74 is consistent with previous findings, it is almost easier to believe an a priori estimate of ES rather than any of the econometric findings. Earlier cross-section estimates (see articles by Stuart Altman and Alan Fechter, Walter Oi, and Altman) were based upon only nine regional observations and had to exclude such plausible enlistment causes as regional preferences, education levels, and variances in the income distribution. The simultaneity between enlistments and youth unemployment was recognized but not avoided. Conversely, Fisher evades simultaneity by lagging his causal variables, and his time-series contains suffi-

* The author is an economist with The RAND Corporation. He is grateful to Steven Canby, Gerti Brunner, and Robert Wilburn for their comments on various ideas contained in this paper. The views expressed are those of the author and do not reflect those of The RAND Corporation or any of its sponsors.

[1] Previous studies by Stuart Altman and Allan Fechter, Walter Oi, and Altman have not theoretically justified their finding of a decreasing elasticity supply curve. Their specific curve may, however, be derived by assuming a Pareto distribution of youth civilian opportunities, though this form would fit poorly the lower tail of the opportunity distribution if the latter were lognormal. Also, Fisher's own approximation equation (1) is more akin to the Pareto than the lognormal distribution.

TABLE 1—IMPACT OF SEASONAL DUMMY VARIABLES IN EQUATION (1)[a]

Coefficients	$ln(W_c/W_{m-1})$	$ln(1-U/P)_{-1}$	$ln(1-A/P)$	R^2	$D.W.$
With Dummies: (1)	$-.007$	$-.009$	$-.312$	.90	.31
	(.003)[b]	(.010)	(.041)		
Without Dummies: (1a)	$-.007$	$-.017$	$-.444$	.70	2.14
	(.005)	(.009)	(.060)		

[a] The wage and unemployment variables are lagged to avoid a simultaneous equation bias, but an unlagged wage coefficient is the same to three decimal places so the lag is not critical. Data are taken from the monograph of John Hause and Fisher.
[b] Standard errors in parentheses.

cient observations. However, we are not told that there is strong serial correlation in the residuals of equation (1) so that the fit is not as good as indicated; there is still a simultaneous equations bias (downward) in ES; and the implied elasticity of enlistments to the unemployment rate is low.

It is well known that the enlistment rate (E/P) varies seasonally, but there may be no need to introduce seasonal dummies into equation (1) because A/P, the draft pressure surrogate, also is seasonal. In fact, the use of seasonal dummies is associated with serially correlated residuals in equation (1), as shown by the Durbin-Watson $(D.W.)$ statistic in Table 1. The wage coefficient remains unchanged, but the unemployment and draft pressure coefficients are increased in absolute value when dummies are dropped. This may explain why Fisher's elasticity of enlistments to the unemployment rate is low.[2] More importantly, it may explain why his draft pressure coefficient is weak.

When Fisher sets draft pressure equal to zero in equation (1),[3] his enlistment rate decreases by only 24 percent. Presumably these are the draft-motivated enlistees, but this 24 percent is much lower than the 38 percent indicated by 1964 attitude surveys of first-term enlistees by the Department of Defense. He gives excellent reasons for the discrepancy,[4] but we have just seen that the seasonal dummies weaken draft pressure. In fact, using (1a) in Table 1 we obtain a reasonable 41 percent for draft-motivated enlistees. Perhaps the exclusion of seasonal dummies spuriously strengthens the draft pressure effect; however, this is offset somewhat by the imperfect definition of draft pressure.

To obtain the true wage, unemployment, and draft effects from (1a), we must correct its coefficients for simultaneous equations bias. The military accession rate A/P is an exogenous variable set by military requirements: but E/P and I/P, the enlistment and induction rates, are endogenous. Furthermore, though not perfect by any means, I/P seems to be a better indicator of draft

[2] In equation (1): $ES(V/P,\ U/P)=ES(V/P,\ 1-U/P)\times ES(1-U/P,\ U/P)=-[.009/(V/P)][-(U/P)/(1-U/P)]=-1(-.18)=.18$. Studies by Altman and Fechter, and Altman indicate a .26 coefficient. Without dummies, however, we derive .30 from (1a) in Table 1 and, allowing for simultaneous equations bias (explained later), we obtain a reasonable .62 elasticity of volunteers to unemployment.

[3] That is, inductions equal zero, so $A/P=E/P$.

[4] Another reason lies in the inevitably imperfect definition of draft pressure which is really a psychological variable. This errors-in-the-variable effect reduces the absolute value of any draft pressure coefficient and hence understates the draft pressure effect. The draft pressure coefficient in equation (1) is $-b_1/(1+b_1)$ because of the identity $A=I+E$, so a 10 percent underestimate in b_1 may cause a 20 percent error in the DPC. See John Johnston, p. 150, and Fisher, p. 250.

Also perhaps the nonrandom residuals in (1) are caused by the omission of an important explanatory variable such as the number of potential and unobservable enlistees turned away from recruiting stations. The non-Army services did turn category III people away during 1957–65 and surely not all enlisted in the Army. In (1) the true enlistment rate should be E^*/P where $E^*=$true number of enlistees. Thus $E^*/P=(E/P)+(E^*-E)/P$ and the observed relation is (1) with the added variable $(E^*-E)/P$ on the right side. Its coefficient is minus one and it should be positively correlated with the included variables. However, (1) excludes $(E^*-E)/P$ so the estimated coefficients of (1) are probably underestimates of the true coefficients. (See Henri Theil, p. 326.) This phenomenon may explain the low response elasticity of E/P to Fisher's variables.

pressure than A/P. We may think of I/P as a proxy for true draft pressure. Thus I/P should replace A/P in equation (1). Using the argument of Fisher's equations (22)–(24) we obtain

$$E/P = b_0 + b_1 \ln(W_c/W_m) + b_2 \ln(1 - U/P) + b_3 \ln(1 - I/P), \tag{2}$$

where

$b_2 = b_1 a_1$ and $b_3 = b_1 a_2$;

a_1 = elasticity of the subjective probability of employment (P_e) to changes in the employment rate;

a_2 = elasticity of the subjective probability of remaining a civilian (P_c) to changes in $(1 - I/P)$.

The presence or absence of seasonal dummies does not affect the following argument.

In a Taylor series expansion, $\ln(1 - I/P) \simeq -I/P$ when $I/P \simeq 0$ so we have an identity.[5]

$$A/P = I/P + E/P \tag{3}$$

Solving equation (3) for $-I/P$, inserting the result into equation (2), rearranging terms and letting $c = 1 - b_1 a_2$ gives

$$E/P = b_{0/c} + (b_1/c) \ln(W_c/W_m) + (b_1 a_1/c) \ln(1 - U/P) - (b_1 a_2/c)(A/P) \tag{4}$$

When we use the results of Table 1, Fisher's estimating equation both with and without seasonal dummies, we obtain

With Dummies	*Without Dummies*
$b_1 = -.010$	$b_1 = -.013$
$a_1 = 1.26$	$a_1 = 2.4$
$a_2 = 43.94$[6]	$a_2 = 63.4$

Now ES in equation (4) is shown by Fisher to be of the declining elasticity form

$$ES = -b_1/(E/P) \tag{5}$$

Since $E/P = .015$ in his quarterly observations and, following the Department of Defense survey we assume 62 percent of E/P are true volunteers, we have $V/P = .009$ so

$$ES = .010/.009 = 1.11 \text{ (with seasonal dummies) or}$$
$$ES = .013/.009 = 1.44 \text{ (without dummies)}^7 \tag{6}$$

Fisher's $ES = .74$ is significantly below these corrected elasticities.

The new derived elasticity of volunteers to unemployment is also reasonable. Previous estimates[8] indicated $ES\ (V/P,\ U/P) = .26$, and equation (1) points to an even lower .18. Fortunately, after adjusting for simultaneous equations bias, we obtain a quite high .62 elasticity. In equation (2), $b_2 = b_1 a_1 = (.-013)(2.4) = -.0312$ so, with $V/P = .009$, $ES((V/P), 1 - U/P) = -.0312/(V/P) = -.0312/.009 = -3.43$. Furthermore, $ES(1 - U/P,\ U/P) = (-U/P)/(1 - U - P) = (-.15)/(1 - .15) = -.18$ so $ES(V/P, U/P) = (3.43)(-.18) = .62$ which is twice as great as the estimates summarized in the Hause and Fisher study.

As for draft pressure, from (2) we have $b_3 = b_1 a_2 = -.013(63.4) = -.824$. But $\ln(1 - I/P) \simeq -I/P = .0077$ so setting draft pressure $I/P = 0$ decreases E/P by $b_3(I/P) = -.24(-.077) = .0063$, a 41 percent reduction from the E/P average of .0153. Thus it is predicted that 41 percent of enlistees were draft induced during 1957–65. This conclusion agrees closely with the 1964 *DOD* survey of enlistee attitudes.

In summary, with seasonal dummy variables excluded and simultaneous equations bias reduced, we obtain acceptable estimates for all supply function coefficients. Con-

[5] This identity ignores the approximately 5,000 category IV enlistments each quarter. In practice, quarterly $I/P = .0077$.

[6] This is $ES(P_c, 1 - I/P)$ but $ES(P_c, I/P) = ES(P_c, 1 - I/P) \times ES(1 - I/P, I/P) = 43.94(-.01) = -.44$ which seems reasonable: a 10 percent rise in draft pressure reduces the subjective probability of remaining a civilian by 4.4 percent.

[7] About 30 percent of P are physically, mentally and morally unfit for service according to Oi, p. 43. But this correction of P does not affect the elasticity estimates.

[8] See Hause and Fisher, (p. 68) for a comparison.

versely, Fisher's draft pressure and unemployment coefficients are smaller:

ELASTICITY COMPARISONS AT THE TRUE VOLUNTEER MARGIN

	Fisher Equation	Amended Version
Wage elasticity	.74	1.44
Unemployment elasticity	.18	.62
Proportion of draft-induced enlistees	.24	.41

These differences dramatically decrease the volunteer cost estimates and imply that the force flexibility is greater under voluntarism than previously imagined. However, the relatively high supply elasticities indicate that recruiting deficits may arise easily if if military salaries are allowed to lag.

II. *The Cost of Voluntarism in 1970*

Fisher calculates the wage increase necessary to provide a 1957–65 voluntary enlistment rate sufficient to stock a 2.65 million-man active duty force in 1965. But this is based on the 1957–65 level of P (male civilians aged 17 to 20 years). Allowing for the 30 percent increase in youth between 1957–65 and 1970 (indicated by the Census Bureau), we may obtain a 2.65 million-man force in 1970 with a V/P 30 percent lower than in 1957–65. With this correction, and assuming no change in military aversion, Fisher's \$4.8 billion cost increment erodes to \$.3 billion, or \$.4 billion allowing for a five-year wage level inflation to 1970. The cost is very low because Fisher's draft pressure coefficient indicates only 24 percent of all enlistees are draft induced.

Since P grows 30 percent the V/P ratio may fall 30 percent from .015 to .012 in 1970, and the V numerator will remain constant and sufficient to stock a 2.65 million-man force. Projected V/P is only .0089, with a draft pressure coefficient of .41 so voluntary enlistments must be raised 33 percent to the $V/P=.012$ target (Fisher, p. 251). Now formula (6) indicates $ES=1.44$ at the mean volunteer rate of .009, but ES should fall somewhat as V/P rises from .009 to .012. The mean elasticity over the range is $ES=1.26$.

With $ES=1.26$, a 33 percent increase in volunteers is obtainable with a 26 percent wage increase. We assume that the average first-term pay in 1970 will be about \$3,500 (includes base pay plus imputed value of food, clothes, shelter and tax advantage), and that the *perceived* value to the serviceman will be about \$2,600.[9] The required base pay increase is only 26 percent of this amount, or \$676 per man per year over his first tour of duty. This is about a third of Fisher's \$2,105 increase, and it does not cause wage inversion. The added wage bill for 1.15 million first-termers is only \$0.78 billion,[10] and this increment ignores the possibility of more intensive use of recruits in mental category IV. With such an adjustment the increment would be almost zero.

Conversely, Fisher omits several important cost items: the differential pay (presumably in bonuses) to overcome aversion to Army life, increases in officer pay, increases in seniority pay from the shifting composition of a volunteer force, and increased retirement costs. Given an Army supply elasticity that is close to unity in the relevant range (see Oi, p. 48), the use of 20 percent lower-quality enlisted manpower, and the lower inflow requirements of a vol-

[9] Enlistees apparently do not perceive the full value of in-kind military services or do not value them highly. Enlistees valued base pay plus supplements at only \$1550 in 1963 according to one unpublished survey by S. Canby, or 65 percent of the military's accounting value of base pay plus the in-kind value of food, clothing and shelter. We assume this perception will rise to \$2600 in 1970.

[10] The estimate assumes youth unemployment rates of 1957–1965 hold in 1970. This seems plausible since the youth pool is growing while demand for unskilled labor seems to be falling. With $ES(V/P, U/P)=.62$, an unlikely 20 percent fall in U/P would cause V/P to drop 12 percent to .008, or 50 percent short of the 1970 target. This would necessitate a 40 percent increase in first-term wages and cause wage inversion in the force. The cost of 40 percent to first-termers is about \$1.2 billion, plus 15 percent to others to avoid aversion equals \$.4 billion, a total of \$1.6 billion. Also, if it is necessary to raise all enlisted men's wages by \$676 (to avoid friction and obtain the full impact on potential careerists) the across-the-board cost is \$676×(2.32 million men) = \$1.59 billion. In this case the wages of first termers should be raised as much as possible to achieve more effect on PPV, the perceived present value of career earnings.

TABLE 2—ADDED COSTS OF VOLUNTARISM
(In Billions)

Enlisted men	\$ .78
Army bonus	.36
Officers	.10
Seniority	1.60
Retirement	.85
Training Savings	1.60
	\$2.09

unteer army, the bonus cost should be no more than \$0.36 billion with a \$2,000 enlistment bonus.[11] Officer requirements of 32,000 may be met by increasing their first-term pay from 16 to 30 percent and giving Army officers an additional \$1,500 to \$4,000 enlistment bonus, depending on the unemployment rate. This costs \$0.03–0.10 billion. Added seniority costs are estimated at \$1.6 billion. Increased retirement costs are due to an increase in the retirement flow and, calculated by *DOD* methods, are about \$0.85 billion. Savings from reduced training activities are probably \$1.6 billion.[12] Thus, surrounded by a penumbra of uncertainty, the total cost increment of voluntarism (Table 2) is only \$2.09 billion and most of this (retirement and seniority) develops in the long run.

If voluntarism is postponed to the mid-1970's, the incremental cost should be approximately the same. Although wages will have to be increased from a larger base and U/P will probably fall due to a deceleration in P, the growth of P should lower required V/P enough to offset these cost increasing effects.[13]

Our cost increment, though considerably below some estimates, is most consistent with the Oi study that puts a \$4-billion upper bound on the cost but excludes the use of mental category IV recruits. In their paper Altman and Fechter estimate a \$5-billion *lower* bound but forget the higher retention rates of voluntarism. Both studies exclude the possibility of an Army bonus and ignore retirement costs. Also, no study (ours included) has estimated the cost of obtaining a voluntary reserve force.

III. *A Theory of the Military Supply Curve*

If the frequency distribution of expected civilian opportunities (and of military aversion) determines the shape of a military supply curve, it is merely the cumulative curve of the frequency function, as Fisher indicates. After making a case for lognormality of civilian opportunities, however, he derives an elasticity formula from an underlying *normal* distribution. The form is quite complicated and has ranges of increasing and decreasing elasticity. The critical point is somewhere in the range of $u/2<X<u$, where u = mean of the normal distribution.[14] Computer evaluation shows X to be closer to u than to $u/2$ for reasonable cases.

Fisher argues that military wages W_m are sufficiently high compared to civilian wages W_c so that $W_m/W_c \simeq 1$ which indicates our observations are on the decreasing elasticity portion of his curve. However, when one considers that a military aversion factor must be added to W_c or subtracted from W_m, assuming average aversion is positive, and that only a portion of military wages are perceived, the wage ratio is certainly less than unity and our observations could easily lie on the increasing elasticity segment of the supply curve. This possibility is a virtual certainty if the frequency curve is normal.

The problem of what relative wages really are is easily avoided by returning to Fisher's original specification of a lognormal curve. The cumulative function of this curve exhibits decreasing elasticity over its *entire*

[11] It is only \$23 million if youth unemployment rates are 14.9 percent. These costs, and those following are based on the study by the author and Steven Canby.

[12] Canby and Klotz, Appendices A and C.

[13] During the 1970s P will grow about 18 percent so that V/P can fall 18 percent and V will not change. But U/P may fall from its 1957–65 average of 15 percent to only 10 percent as P decelerates. With $ES(V/P, U/P)$ $=.62$, cost forces should remain balanced throughout the 1970s.

[14] Fisher, p. 253. This behavior is simulated by the logistic function $y/(1-y)=e^{a+bX}$, which has $ES = bX(1-y)$.

range. That is, using Fisher's procedure, define

(7) $$x = ln(W_m)$$

so the cumulative normal distribution function equals the voluntary enlistment rate $y=(V/P)$:

(8) $$y = (2\pi\sigma^2)^{-1/2}\int_{-\infty}^{x} e^{-1/2((t-u)/\sigma)^2}dt,$$

where u = mean and σ = variance of the normal distribution of the logarithm of civilian opportunities (plus military aversion). Now

(9) $$ES(y, x) = (x/y)(dy/dx)$$

but from (7),

(10) $$dx = dw/w$$

so putting (10) into (9) we obtain

(11) $$\begin{aligned} ES(y, x) &= x(w/y)(dy/dw) \\ &= x \times ES(y, w) \end{aligned}$$

Thus

(12) $$ES(y, w) = ES(y, x)/x = (dy/dx)/y$$

Also, from equation (8),

(13) $$dy/dx = (2\pi\sigma^2)^{-1/2}e^{-1/2((x-u)/\sigma)}$$

Matching normal curve tables of marginal (dy/dx) and cumulative (y) frequencies reveals a mildly declining ES as y rises.

We may now construct a theoretical, or predicted, elasticity. From (6) we have quarterly $V/P=.01$ so annual $V/P=.04$, which is the annual voluntary enlistment rate y to be put into equation (12) to obtain ES at the volunteer margin. When $y=.04$, however, the normal curve frequency table indicates $x=u-1.75\sigma$. Inserting this value into equation (13) and calculating $(2\pi)^{-1/2}=.399$ we obtain

$$dy/dx = .086/\sigma$$

Now from equation (12) we have

(14) $$ES = .086/.04\sigma = 2.15/\sigma$$

Thus, given the volunteer enlistment rate, $y=.04$, the supply elasticity of recruits depends inversely upon σ.[15] Unfortunately, σ is unknown and we can only make plausible guesses about it derived from guesses about the mean and variance of the assumed lognormal distribution. Evidently the mean and standard deviation of young men's earnings are \$3500 and \$1071, respectively.[16] If we arbitrarily add \$500 to the mean and \$200 to the standard deviation for military aversion, the ratio of the standard deviation to the mean $(\sigma_w/\bar{w})$ becomes .33. Using a standard text[17] we may find the corresponding σ for the normal distribution of $ln\ w$ as

(15) $$e^{\sigma^2} = 1 + (\sigma_w/\bar{w})^2 = 1.11,$$

so that $\sigma^2=.095$ and $\sigma=.31$. Thus, from equation (14), the theoretical $ES=6.96$, which is vastly greater than Fisher's .74.[18]

The discrepancy is resolved by reorienting the theory from a wage elasticity to a perceived present value elasticity. In fact, Fisher's theory applies only to *careerist* volunteers. To prove this, redefine

(16) $$X = ln\ PPV,$$

where PPV is the perceived present value of career earnings. We may think of

$$PPV = w\sum_{i=1}^{40}[(1+r)^i/(1+d)^i],$$

where w = initial salary, r = rate of expected wage increase, and d = subjective discount rate. The unknown mean, variance, and correlation of these three elements in the population determines the frequency distribution of PPV. If the population pool P has a lognormal distribution of civilian PPV, then the argument of equations (7) through (14) holds for PPV also. However, σ^2 now refers to the variance of the logarithm of PPV, not of wages. If initial

[15] As indicated above, equation (14) also declines for larger y: if $y=.05$, then $ES=2.04/\sigma$.

[16] Fisher, p. 253, fn. 19.

[17] Maurice G. Kendall and Alan S. Stuart, p. 68.

[18] Our results are sensitive to different aversion correction factors but ES remains high. For example, if the ratio of standard deviation to mean is as high as .7, then $\sigma=.65$ and $ES=3.38$.

$w=\$4000$ (including aversion), $r=5$ percent and $d=10$ percent, then $PPV \simeq \$80{,}000$ for a 40-year career. The σ of PPV should be relatively larger than the standard deviation of one year's expected earnings because the variance of a sum equals the sum of the component variances if the components are independent. This independence is not strictly true of year to year income, but a tendency should exist to increase the variance. If $(\sigma_w/\bar{w})=.33$, we expect $\sigma_{PPV}/80{,}000 > .33$. The extreme skewness of PPV due to the compounding of different r and d values, and a certain additivity of variances, suggests $\sigma_{PPV}/PPV=.5$ is possible so that $\sigma_{PPV}=\$40{,}000$. Now equation (15) indicates $\sigma=.49$ and equation (14) indicates $ES=4.57$.

Such an ES greatly exceeds unity for reasonable y and σ values, but remember it is $ES(y, PPV)$, not $ES(y, W_m)$. We need another elasticity element, $ES(PPV, W_m)$, to close the chain. Defining W_m now as an average (or across the board) wage variable, this elasticity is unity only for volunteers who plan on a military career; a doubling of W_m doubles their PPV_m. But unpublished Air Force surveys indicate that only about 16 percent of all true volunteers in 1964 considered a service career at the time of entry; all others were undecided or short-termers of various lengths who were seeking training or a place to mature. These young men also have a high subjective discount rate. A doubling of W_m, then, probably increases the PPV_m of short-termers by about 15 percent, assuming they stay for one four-year term (one-tenth of their earning life) and have a 10 percent discount rate of earnings. Thus, perhaps $ES(PPV_m, W_m)=.15$ for this short-term group.

The aggregate $ES(PPV, W_m)$ is then a weighted average of the elasticities of careerist and noncareerist volunteers; the weight is the proportion (K) of true volunteers who are intended careerists. Obviously this proportion is a function of relative wages, since everyone would be a careerist at very high W_m. At current wage relatives, however, we assume $K=.16$. Since the careerist elasticity (ES_c) is unity and the noncareerist elasticity (ES_{nc}) about .15, we have a crude estimate of

$$\begin{aligned} ES(PPV_m, W_m) &= ES_c(K)+(1+K)ES_{nc} \\ &= 1(.16)+(1-.16)(.15) \qquad (17)\\ &= .29 \end{aligned}$$

Now we may obtain an a priori estimate of

$$\begin{aligned} ES(y, W_m) &= ES(y, PPV_m) \\ &\quad \cdot ES(PPV_m, W_m) \qquad (18)\\ &= 4.57(.29) = 1.32 \end{aligned}$$

which is about twice Fisher's elasticity estimate for a draft-free world and about equal to our $ES=1.44$ of the first section.

The $ES=1.32$ guesstimate is sensitive to the unknown σ_{PPV} parameter. Thus, if $\sigma_{PPV}=\$80{,}000$ so that $\sigma_{PPV}/\overline{PPV}=1$ we derive an elasticity less than unity. But this is an extreme case so that $ES>1$ is theoretically indicated for most plausible situations.

Since the noncareerist volunteer has a much lower $ES(y, W_m)$ than his careerist companion, and thus apparently depresses the average $ES(y, W_m)$ to the neighborhood above unity, noncareerists deserve a key role in any theory of military supply. Fisher's theory implicitly forces a man to choose between a military and a civilian career, but he has a third choice: he may enter the service, acquire skill to enhance his civilian opportunities, and then leave. Each young man then faces a triplet, PPV_m, PPV_c, and PPV_{mc}. He enters the service only if PPV_c is the worst of three choices. There is a trivariate distribution of these elements in the population and the number of true volunteers (careerists and noncareerists) may be expressed as

$$\begin{aligned} y &= \int^{R}\!\!\int F(PPV_c, PPV_{mc})\, dF \qquad (19)\\ &= g(PPV_m), \end{aligned}$$

where F and g are functions and $R=$region in which PPV_m or PPV_{mc} exceeds PPV_c.[19]

[19] The conditional distribution $y=g(PPV_m)$ is extremely complicated and depends on the covariance matrix of F. See the article by G. M. Tallis for the theory where F is simply bivariate normal.

The three elements are of course strongly correlated.

We wish to estimate $ES(y, W_m) = ES(y, PPV_m) \cdot ES(PPV_m, W_m)$ but, while changes in W_m cause equiproportionate changes in PPV_m they also cause fractional changes in PPV_{mc}, so equation (19) shifts with W_m. These effects are illustrated in Figure 1.

The population is arrayed in a strongly correlated bivariate distribution of PPV_c and PPV_{mc}. At a given PPV_m the people in region A maximize their PPV in a civilian career; D people find PPV_{mc} a maximum in their choice triplet (the noncareer true volunteers); and B and C people maximize by choosing PPV_m (the careerist volunteers). As indicated before, suppose the population is proportioned $A=96$ percent, $D=3.5$ percent, and $B+C=0.65$ percent. Thus about 16 percent of all volunteers are careerists. Let W_m rise 20 percent so that PPV_m wedge moves outward 20 percent and $B+C$ expands at the expense of A and D. The $B+C+D=V=$true volunteer area unambiguously expands. However, the W_m increase causes PPV_{mc} to rise also and the distribution F shifts rightward by, say, 3 percent.[20]

With this beginning, Appendix A generates $ES(y, W_m)=1.45$ which is similar to our previous guesstimate of 1.32 in equation (18) and equal to our econometric $ES=1.44$ in (6). Our last $ES=1.44$ estimate is then seemingly consistent with the foregoing theories that imply a supply elasticity exceeding unity in the ranges necessary to stock a 2.65 million volunteer force in 1970.

Appendix A

A Guesstimate of the Theoretical Supply Elasticity

In Figure 1 the population is proportioned $A=.96$, $D=.0335$, $B+C=.0065$. Now let W_m rise 20 percent so the PPV_m wedge moves outward 20 percent and $B+C$ expands at the expense of A and D. The $B+C+D=$true volunteer area, then, unambiguously expands. The W_m rise, however,

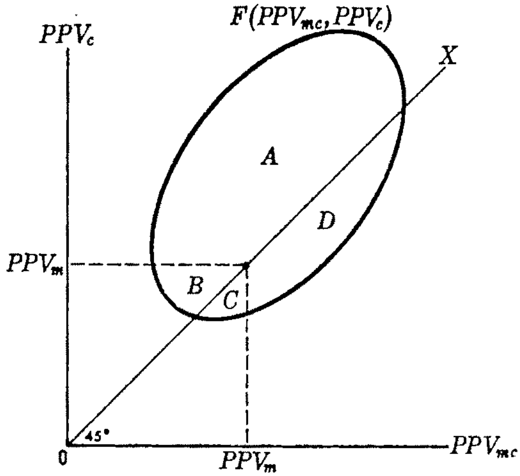


FIGURE 1—CONTOUR OF $F(PPV_{mc}, PPV_c)$

causes PPV_{mc} to increase, and the distribution F shifts to the right by, say, 3 percent. This is equivalent to a leftward displacement of OX by 3 percent.

Previously we had $D/A=.0335/.96 \simeq .03$. Now *if F is approximately normal*, OX must cut F into D and A at a point $u+1.9\sigma$, i.e., $F(x=u+1.9\sigma)=.97$, since .03 of the probability lies beyond $x=u+1.9$ under the normal curve. Assume further that $u=3\sigma$, so $x=4.9\sigma$. Now a 3 percent decrease in x (shift in OX) gives an $x=4.75\sigma$ and $F(u+1.75\sigma)=A=.96$. Thus the percentage change in F, dF/F, equals $dA/A \cong .01/.97 \cong 1\%$. Since $ES(A, W_m)=(dA/A)/(dW/W)$, we have $ES(A, W_m)=1/20=.05$.

Also, above we have $B/A \simeq .0065/.96 \simeq .006$. Thus PPV_m must cut F on the average at a point $x=u-2.52\sigma$. Again, assuming normality, $F(x=u-2.52\sigma)=.006$ or $A=F(x+2.52\sigma)=.994$. With $u=3\sigma$ as before, $x=3\sigma-2.52\sigma=.48\sigma$. Now $x=PPV_m$; so $dW/W=20\%$ means $d(PPV_m)/PPV_m=dx/x=20\%$ also. Thus x rises from $.48\sigma$ to $.58\sigma$, or $x=3\sigma-2.42\sigma=u-2.42\sigma$. Now $F(u-2.42\sigma)=.008$ so $A=F(u+2.42\sigma)=.992$. Thus $dA/A=(.994-.992)/.990 \simeq .2\%$ and $(dA/A)/dW/W)=.2\%/20\%=.01$. The total dA/A is then the sum of the two effects: $1\%+.2\%=1.2\%$. Since $B+C+D=V=y=1-A$, it follows that $dy/y=dV/V=(dA/A)=1.2\%(.94/.04)=29\%$. Since initially $W=20\%$, we have $ES(y, W_m)=29\%/20\%=1.45$.

REFERENCES

S. H. Altman, "Earnings, Unemployment and the Supply of Enlisted Volunteers," *J. Hum. Resources*, winter 1969, *4*, 38–59.

[20] We previously assumed $ES(PPV_m, W_m)$ for noncareerists was .15, so .15(20%)=3%.

——— and A. E. Fechter, "The Supply of Military Personnel in the Absence of a Draft," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 19–31.

S. L. Canby and B. P. Klotz, "The Budget Cost of a Volunteer Military," RM-6184-PR, Santa Monica 1969.

R. L. Chaney, *Discounting by Military Personnel by Various Ages*, unpublished Department of Defense Study, Washington, October 1962.

A. C. Fisher, "The Cost of the Draft and the Cost of Ending the Draft," *Amer. Econ. Rev.*, June 1969, *59*, 239–54.

J. C. Hause and A. C. Fisher, *The Supply of First-Term Enlisted Manpower in the Absence of a Draft*, Institute for Defense Analyses, Study S-293, Washington, Apr. 1968.

J. Johnston, *Econometric Methods*, New York 1963.

M. G. Kendall and A. S. Stuart, *The Advanced Theory of Statistics*, Vol. 2, New York 1961.

W. Y. Oi, "The Economic Cost of the Draft," *Amer. Econ. Rev., Proc.*, May 1967, *57*, 39–62.

G. M. Tallis, "Plane Truncation in Normal Populations," *J. Roy. Statist. Soc.*, Series B, 1965, *27*, 301–307.

H. Theil, *Economic Forecasts and Policy*, Amsterdam 1961.

The Cost of Ending the Draft: Reply

By ANTHONY C. FISHER*

Benjamin Klotz's comment makes a contribution to both the theoretical and empirical study of military manpower supply—although I take issue with some of his specific criticism. In what follows I will indicate where I agree, extend the area of agreement, and show where and why I disagree.

I. *The Theory of Military Manpower Supply*

Let us consider first the theory. Here I find a fruitful complementarity. Adopting the model set forth in my equations (1)–(25), Klotz presents in Section IV an argument based on my derivation (pp. 252–53) of the behavior of the elasticity of supply, that elasticity decreases over the entire range of increasing enlistment, rather than first increasing and then decreasing with increasing military wage offers. He then derives an a priori estimate, consistent with an amendment of my empirical estimate, of the decreasing elasticity in the range necessary to stock a 2.65 million-man force. This is an interesting exercise, and I believe a real contribution to the theory of military manpower supply. Further, the initial argument that a first increasing then decreasing elasticity is generated by a normal, not lognormal, distribution of expected civilian alternatives (and preferences) is correct. Although I don't agree with the suggestion that our observations (and still less our projected observations) are likely to lie on the increasing elasticity segment of such a supply curve,[1] the problem is, as Klotz points out, easily avoided by returning to the original specification of an underlying lognormal distribution of alternatives (and preferences). This I now do, to present a rigorous proof of the decreasing elasticity property of supply.

As before, let us define elasticity

$$E(y, x) = \frac{dy}{dx}\cdot\frac{x}{y} \tag{1}$$

where $x = \ln W_M$

$$y = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x} e^{-\frac{1}{2}((t-\mu)/\sigma)^2}dt$$

$$\frac{\frac{dy}{dw}}{w}\cdot\frac{x}{y} \tag{2}$$

where $w = W_M$ and $\frac{dx}{dw} = \frac{1}{w}$

Then

$$E(y, w) = E(y, x)/x = \frac{dy}{dx}\Big/ y \tag{3}$$

Note the difference between $E(y, w)$ in (3) and $E(y, x)$ in (1). In the earlier proof I worked, mistakenly, with $E(y, x)$, deriving behavior of elasticity along a normal, not lognormal, curve. The problem is now to investigate the effect on $E(y, w)$ of changes in x. By differentiation

$$\frac{dE(y, w)}{dx} = \frac{y\frac{d^2y}{dx^2} - \left(\frac{dy}{dx}\right)^2}{y^2} \tag{4}$$

$$\text{num}\,\frac{dE(y, w)}{dx} = \tag{5}$$

$$\left(c\int_{-\infty}^{x} e^{-\frac{1}{2}((t-\mu)/\sigma)^2}dt\right)\cdot\left(ce^{-\frac{1}{2}((x-\mu)/\sigma)^2}\left(\frac{\mu - x}{\sigma^2}\right)\right) - \left(c^2e^{-((x-\mu)/\sigma)^2}\right)$$

* Assistant professor of economics, Brown University.

[1] Reference to fnn. 13 and 14 (p. 246) indicates that military earnings are high enough and preferences for at least sampling military service widespread enough, for our observations to lie on the decreasing elasticity segment. This is still more likely to be true as military earnings are increased to attract additional volunteers.

where $c = \frac{1}{\sqrt{2\pi\sigma^2}}$

$$(6)\quad \text{num}\,\frac{dE(y, w)}{dx} = \left(\int_{-\infty}^{x} e^{-\frac{1}{2}((t-\mu)/\sigma)^2}dt\right)\cdot\left(e^{-\frac{1}{2}((x-\mu)/\sigma)^2}\left(\frac{\mu - x}{\sigma^2}\right)\right) - e^{-((x-\mu)/\sigma)^2}$$

In order to determine the sign of this expression we ask for what values of x, if any, is

$$(7)\quad \left(\frac{\mu - x}{\sigma^2}\right)\int_{-\infty}^{x} e^{-\frac{1}{2}((t-\mu)/\sigma)^2}dt \geq e^{-\frac{1}{2}((x-\mu)/\sigma)^2}$$

This is obviously impossible for $x \geq \mu$, so let us investigate $x < \mu$.

Let $z = \frac{\mu - x}{\sigma}$ (hence $z > 0$)

Then (7) becomes

$$(8)\quad \frac{1}{\sigma}\int_{-\infty}^{\mu-\sigma z} e^{-\frac{1}{2}((t-\mu)/\sigma)^2}dt \overset{?}{\geq} \frac{1}{z}e^{-\frac{1}{2}z^2}$$

Let $S = \frac{\mu - t}{\sigma}$

Then (8) becomes

$$(9)\quad \frac{1}{\sigma}\int_{+\infty}^{z} e^{-\frac{1}{2}s^2}(-\sigma ds) \overset{?}{\geq} \frac{1}{z}e^{-\frac{1}{2}z^2}$$

$$(10)\quad = \int_{z}^{+\infty} e^{-\frac{1}{2}s^2}ds \overset{?}{\geq} \frac{1}{z}e^{-\frac{1}{2}z^2}$$

But[2]

$$(11)\quad \int_{z}^{\infty} e^{-\frac{1}{2}s^2}ds < \frac{1}{z}e^{-\frac{1}{2}z^2} \quad \text{for all } z > 0$$

[2] For proof, see William Feller (p. 175). I am indebted to Professor Walter Strauss of the Brown University mathematics department for calling my attention to this result.

Hence $dE(y, w)/dx$ is *never* >0, and $E(y, w)$ is decreasing over the entire range of x.

II. *Empirical Estimation of the Supply Curve*

Turning to the empirical work, Klotz points out that there is no need to introduce seasonal dummies into the regression equation (equation (25) in my article), because the accession rate A/P, the draft pressure surrogate, is seasonal. Further, introducing seasonal dummies seems to generate serial correlation. This is an insight I am happy to acknowledge, and I accept the amended equation, dummies excluded, with the estimated coefficients given in Klotz's Table 1, equation(1a). It is further reassuring to find the proportion of draft-induced enlistees increased from 24 to 41 percent, consistent with previous survey results.

I do not however accept the assertion, which springs perhaps from a misreading (for which I may be partly to blame) that the amended coefficients are biased. I will show that the coefficients are not biased, that the only relevant criticism of the empirical work is that acknowledged above, and that even it is less important than is suggested in the comment.

Klotz notes that the accession rate, A/P, is an exogenous variable, and then proceeds to replace it with the induction rate, I/P, clearly an endogenous variable. This is especially puzzling in view of my discussion (p. 245–46) of this same point, concluding that, in order to estimate the single supply equation parameters without bias, it is necessary that all independent variables be exogenous, hence the specification of A/P, not I/P. Klotz's equation (2), containing I/P as an independent variable, clearly cannot be estimated without bias. Equation (4), specifying A/P, can, however, be estimated, and this is very nearly what I have done. Equation (35) is the (unbiased) estimated equation, given as equation (1) in the comment, and from it is derived equation (38), used to calculate the effect of the draft, and later of changes in military earnings, on enlistments (p. 251). Equation (38) can be written as

$$\frac{E}{P} = b_0 + b_1 \ln\left(\frac{\tilde{W}_c}{W_M}\right) + b_2 \ln(1 - U) - b_3\left(\frac{I}{P}\right) \tag{12}$$

where $b_1 = b_1'/1 + b_2'$ with $b_1' = -.007$ and $b_2' = -.312$ the estimated coefficients. In a table of coefficients calculated from these estimates following his equation (4), Klotz gives a value for b_1, here presumably the true structural coefficient relating enlistments to earnings, of $-.010$. This is just the value calculated from equation (38), or (12) above, with $b_1 = -.007/1 - .312 \cong -.010$. Further, earnings elasticity at the mean enlistment rate (without a draft) is given by my equation (28) as $E = b_1(E/P) = -.010/.009 \cong -1.11$, where the mean enlistment rate *with* a draft is about .015 (quarterly), and about 60 percent of enlistees are true volunteers (from the *DOD* survey), leaving a mean enlistment rate without a draft of about .009. This is, of course, Klotz's result (equation (6)), and here I am perhaps partly to blame for the seeming confusion, not having *explicitly* re-calculated elasticities after deriving equation (38) from the estimating equation. It is, however, clearly stated (p. 251) that equation (38), and implicitly the associated elasticities, are used to calculate the increase in military earnings necessary to generate a volunteer enlistment rate sufficient to maintain desired levels of armed forces strength.

Removing seasonal dummies raises the absolute value of coefficient and elasticity to $b_1 = .013$, and $E = .013/.009 = 1.44$, respectively, as shown also by Klotz. Again, I accept this amendment, but note at the same time that its quantitative significance is not as great as suggested in the comment.

Klotz's summary statement, "... with seasonal dummy variables excluded, and simultaneous equations bias reduced, we obtain acceptable estimates ..." (p. 972) is, then, unacceptable. I have shown that the only difference in estimates, and hence in elasticities, is that due to the exclusion of seasonal dummies. Klotz concludes this section of his comment by asserting that "these differences [in parameter estimates] dramatically decrease the volunteer cost estimates." No evidence to support this assertion is forthcoming, though he may be looking ahead to 1970. I will have some comments on "the cost of voluntarism in 1970," but first, let us look into the effect on 1965 volunteer cost estimates of the amended coefficients.

Recall that the mean enlistment rate is .015, which falls to .009 without the draft-induced enlistments (40 percent of the total). The desired enlistment rate, E^*/P, is again the accession rate adjusted by a factor of .707 (p. 251), hence $E^*/P = .707\ (.023) \cong .016$. Thus the enlistment rate must be increased from .009 to .016, or by about 78 percent. Elasticity varies from $E = .013/.009 \cong 1.44$ to $E = .013/.016 \cong .81$ over this range, with a mean of about 1.12 which, applied to the desired increase in enlistments, yields an increase in military earnings of approximately 70 percent, hardly a "dramatic change" from the 80 percent associated with the original coefficients.[3]

III. *The Cost of Ending the Draft in 1970*

Now, what can be said about the cost of ending the draft in 1970? This is a question not considered in my article, but as Mr. Klotz uses my model and amended estimates in his discussion, I will venture a few remarks.

Accepting the rough calculations yielding a desired enlistment rate of $E^*/P = .012$, the indicated 33 percent increase (from .009) in the enlistment rate requires a 26 percent increase (the mean elasticity over this range is 1.26) in military earnings. This is of course an increase in military earnings *relative* to civilian, as intuition and model alike tell us that increasing military earnings will not induce additional enlistments if civilian earnings are similarly increasing (other things equal). Klotz calculates the required (26 percent) increase, and applies it to assumed first-term military pay in 1970 to derive the budgetary cost of attracting a volunteer

[3] The change is even less than the 10 percentage points indicated, rounding errors causing the amended increase (70 percent) to appear lower than it really is.

force in 1970. Nothing is said about civilian earnings, so the implicit assumption is that they are the same, relative to military, as in 1965. This may be a good assumption, but I should like to see some evidence for it, including a discussion of the effects of inflation over this period.

Further, it is not clear that other things, namely youth unemployment and the distribution of tastes for military service, have remained the same. For example, while the unemployment rate for 18–19 year-old males averaged about 15 percent over the period 1957–65, it dropped to 10 percent in late 1965, and has continued through 1969 to fluctuate around that level. Klotz notes the possibility that a falling unemployment rate could cause an increase in the military wage offer required to attract a volunteer force, but assumes that youth unemployment rates of 1957–65 will hold in 1970. In view of the evidence, this seems a curious assumption. Thus a 33 percent fall in unemployment, from 15 percent to 10 percent would, given the unemployment elasticity of .62, result in a 20 percent fall in enlistments, necessitating a total increase of 46 percent in military earnings. Further, an increase of this magnitude to first-termers would, unlike the 26 percent increase, cause pay inversion, necessitating additional pay to non-first termers. In defense of Klotz' approach, it might be conjectured that as the fall in unemployment was associated with a buildup in armed forces, return to a 2.65 million man force could result in a rise in unemployment, perhaps to 1957–65 levels. My own feeling is that this is not likely for several reasons, including effects of the new G.I. Bill in removing young men from the labor force. It is a debatable point, and deserves careful consideration in any attempt to predict the budgetary cost of a volunteer force in 1970 and beyond.

Another problem for prediction, noted above, is uncertainty about the distribution of tastes for military service in the wake of national and international events of the past five years. My casual impression is that the distribution has shifted in a way unfavorable to the military, but even if this is true, return to a smaller force, accompanied by an end to the war in Vietnam, could shift it back. Klotz finds several "important cost items" omitted from my calculation of the budgetary cost of a volunteer force, including "the differential pay (presumably in bonuses) that must be paid to overcome aversion to Army life." But I have cited (admittedly crude) evidence that, on the volunteer margin, potential volunteers are indifferent to the prospect of a military career (p. 242). Besides, why the *ad hoc* introduction of bonuses? Our estimated supply equation tells us what level of military compensation, in the form of bonuses or otherwise, would be required to attract any desired number of volunteers. The real question is, have tastes changed enough to shift significantly the distribution of civilian alternatives and preferences, a part of which (the enlistment rate) is cut off by the military wage offer? I have suggested that we do not know the answer to this question, at least in the event of a return to the conditions (smaller force, no major conflict) underlying the original estimation. Klotz is not explicit here, but his assertion that volunteers must be offered bonuses in addition to the calculated wage increase suggests he believes that preferences are and will remain significantly altered. Once again, this is at best an uncertain proposition, which must be given explicit consideration in any prediction of cost in 1970.

The next "omitted" cost item is "increases in officer pay," which Klotz puts at $.10 billion. Yet I have clearly (or so I thought) stated (p. 252), "Fechter estimates the additional cost of attracting officers in the absence of a draft as $.42 billion," referring to results shown by Stuart Altman and Alan Fechter (p. 3) adopted in my cost calculations (p. 252). Increases in officer pay are, then, not omitted, and I can only suppose that Klotz has simply overlooked their mention. Again, the question arises as to whether earlier results are valid, and here Klotz, with his lower estimate, implicitly suggests they are not. What can we say on this question? The population pool has increased making it easier (less expensive) to

attract a given number of officers. In addition, beginning officer pay has increased relative to civilian alternatives and the shift in preferences (away from the military) is probably more significant in the population of potential officers (college students and college graduates) than in the population of potential enlisted men. The effect of these changes is not clear, a priori, but in view of their offsetting nature I would be inclined to stick with the earlier results.

Summing up the discussion of the cost of voluntarism in 1970: (1) the added wage bill to enlisted men, on the basis of a 46 percent increase in Klotz' projected perceived military earnings to first termers, and *no* increase to non-first termers to avoid inversion, is \$1.38 billion; (2) the added wage bill to officers is \$.42 billion; (3) no *additional* bonus to enlisted men is required; and (4) seniority and retirement cost increases are probably offset by training, efficiency and tax savings. The additional budgetary cost of a volunteer force in 1970 then comes to \$1.8 billion, not very different from Klotz' \$2.09 billion. As to differences in the separate components of cost, I have tried to consider and evaluate these explicitly, and the reader may take his choice. Finally, a caution concerning the uncertainty still surrounding these calculations might be registered.

IV. *Conclusion*

Klotz has, within the framework of my model, provided additional insight into the theory of military manpower supply. In Section I above, I take the analysis a step further, proving that supply elasticity is decreasing over the entire range of increasing military wage offers.

There is an error in his assertion of bias in my time-series estimates. On the other hand, he is right in excluding seasonal dummies from the regression equation.

The discussion of the cost of voluntarism in 1970 seems to me seriously weakened by an apparent ignorance or disregard of such important determinants of volunteer supply as civilian earnings, unemployment, and the distribution of tastes for military service. Klotz makes in effect a number of implicit assumptions about the time-paths of these variables, all questionable at best. In Section III above I bring out these assumptions, show their effect on cost calculations, and suggest hopefully more plausible alternatives. Other cost increments of a volunteer force, notably increases in officer pay, are discussed and evaluated. Subject to the qualification of massive uncertainty, the extra budgetary cost of a 2.65 million-man volunteer force in 1970 would appear to be in the neighborhood of \$1.8 billion (my estimate) to \$2.09 billion (Klotz's estimate).

REFERENCES

S. H. Altman and A. E. Fechter, "The Supply of Military Personnel in the Absence of a Draft," *Amer. Econ. Rev., Proc.,* May 1967, *57*, 19–31.

W. Feller, *An Introduction to Probability Theory and its Applications,* Vol. 1, New York 1968.

A. C. Fisher, "The Cost of the Draft and the Cost of Ending the Draft," *Amer. Econ. Rev.,* June 1969, *59*, 239–54.

International Capital Movements and Monetary Equilibrium: Comment

By H. Robert Heller*

In a recent article in this *Review*, John E. Floyd argues that international "portfolio adjustments . . . financed by once and for all changes in the stock of foreign exchange reserves . . . involve an equivalent transfer of wealth from the surplus country to the deficit country" (p. 486). This, he maintains, is due to the fact that a surplus country will give up income-producing assets in exchange for "unproductive" foreign exchange reserves. Furthermore, he asserts that "it is apparent that foreign exchange reserves are part of wealth, but they are not equivalent in *value* to an equivalent *quantity* of assets in the form of other types of capital. In part, this depends on how much of the foreign exchange reserves are held in the form of currency and gold, and how much are held in the form of income-producing assets abroad" (fn. 19, italics added).

Evidently, Floyd will count foreign exchange reserves which offer an explicit return in the form of interest as part of national wealth, while assigning a zero value to non-interest yielding reserves like gold or currency. Foreign exchange reserves—in the form of gold, foreign currencies, or interest yielding foreign assets—are a class of assets held by the governmental authorities which are highly liquid and whose value is expected to be very stable in terms of foreign units of account. Foreign exchange reserves derive their usefulness, and thereby their value, from the *services* which they are able to provide to a country faced with a balance of payments deficit. These services consist of enabling the authorities to *finance* external disequilibria rather than being forced to undertake costly *adjustments* of the domestic economy to each and every disturbance in the external accounts.

Thus, rather than having a zero rate of return—as implied by Floyd for all non-interest yielding reserves—the implicit rate of return earned on foreign exchange reserves consists of the adjustment costs which would have to be incurred if no reserves were available for financing purposes. Foreign exchange reserves *do* yield a positive rate of return and thereby are part of national wealth. The fraction of the total national wealth portfolio which should be held in the form of foreign exchange reserves is determined by the condition that the expected marginal rate of return on reserves is equal to the marginal rate of return obtainable by holding other assets.

Contrary to Floyd's assertions, the financing of a balance of payments disequilibrium by the accumulation or depletion of foreign exchange reserve inventories involves *no* transfer of wealth but merely represents a reshuffling of the total wealth portfolio of a country: the surplus country acquires reserves and gives up other assets, while the deficit country reduces its stock of reserves while increasing its holdings of other assets.

* Associate professor of economics, University of California, Los Angeles.

REFERENCE

J. E. Floyd, "International Capital Movements and Monetary Equilibrium," *Amer. Econ. Rev.*, Sept. 1969, *59*, 472–92.

International Capital Movements and Monetary Equilibrium: Reply

By John E. Floyd*

In his comment on my recent article, H. Robert Heller has focused attention on an important issue—what is the effect on the level of wealth of a sale of capital assets in return for foreign exchange reserves, or more generally, of an increase in the stock of foreign exchange reserves from whatever source? In my paper I assumed for analytical convenience that foreign exchange reserves held by the government are not regarded by individuals in the community as part of their wealth, and then qualified this statement in a footnote, arguing in effect that an accumulation of foreign exchange reserves in exchange for capital assets would reduce wealth, but not by the full amount of the reduction in capital assets. Even if they yield no return in the form of interest, such foreign exchange reserves are wealth in the sense that they can be cashed in for real goods or capital assets. If they bear interest, so much the better. This is consistent with the notion, common among trade economists, that surplus countries are giving subsidized loans to deficit countries. The element of subsidy or wealth transfer arises because the reserves are being held in a different form than individuals would choose to hold them, that is, if we were to let the exchange rate float and hand over the stock of foreign exchange reserves on some basis to the members of the community, they would convert them into capital assets at least in part.

In his comment, Heller appears to be making three points. First, he argues that government held foreign exchange reserves have a liquidity value in that they enable the authorities to finance balance of payments deficits without undertaking costly adjustments of the domestic economy. Second, he points out that the fraction of the total national wealth portfolio that should be held in the form of foreign exchange reserves is determined by the condition that the marginal liquidity return noted above be equal to the marginal rate of return obtainable on other assets. Third, he concludes that the financing of a balance-of-payments disequilibrium by an accumulation or depletion of foreign exchange reserves involves no transfer of wealth but merely an exchange of assets. In order to respond effectively to these points, I find it useful to develop more fully the theory of the optimal stock of foreign exchange reserves.

It is possible, as Heller's argument implies, that some or all members of the community may obtain utility from having some degree of exchange rigidity and some government held inventory of foreign exchange reserves which would provide the liquidity necessary to maintain the desired degree of inflexibility of the exchange rate without requiring inflation or deflation of the domestic economy. Since the *raison d'etre* of any government policy is that individuals benefit, the analysis must begin at the individual level. Because individuals differ in their economic circumstances and their attitudes toward risk, they will be affected differently by exchange and domestic income and price variability and the degree of uncertainty with respect to the ability of the government to avoid them. The optimal government held stock of foreign exchange reserves will therefore be different for different individuals. In Figure 1, the vertical axis measures the net increment to wealth resulting from the government held inventory of foreign exchange reserves while the horizontal axis measures the mean or permanent size of that inventory. It is the average or permanent stock of foreign exchange reserves rather than the current stock which is relevant, since the function of the stock of reserves is

* Professor of economics, University of Toronto.

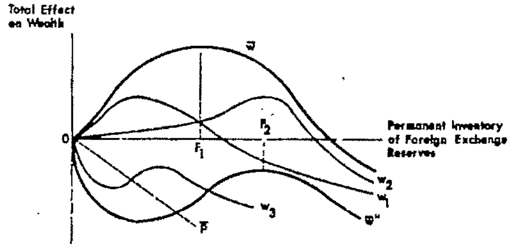


FIGURE 1

to vary randomly in order that other variables in the economy need not vary randomly. I am assuming that a zero mean inventory level implies a flexible exchange rate system. The curves W_1, W_2, and W_3 measure the effects of the maintenance of permanent inventories of various sizes on the wealth positions of individuals 1, 2, and 3. Individual 3, for example, will achieve his maximum wealth position with a flexible exchange rate and a zero inventory of foreign exchange. His wealth curve is drawn to reflect an assumption that once a flexible exchange rate is effectively abandoned, the costs of exchange rigidity can be minimized by holding some positive permanent stock of foreign exchange reserves.

The effect of various inventory levels on the aggregate wealth of the community can be obtained by vertically summing the wealth effects on all individuals. The curves $\overline{W}$ and $\overline{W}''$ give these aggregate wealth effects for all individuals $(1 \ldots n)$ under alternative assumptions, of course, as to the wealth effects on individuals $4 \ldots n$. In the case of $\overline{W}$, the wealth maximizing inventory is F_1: in the case of $\overline{W}''$, a zero inventory (and the associated flexible exchange rate) maximizes aggregate wealth, although once a significant degree of exchange rigidity is established a "second-best" maximum occurs at F_2. The inventory that maximizes aggregate wealth is also the socially optimal one in the sense that it is possible for society, through appropriate income redistributions, to make everyone better off at that inventory than at any other. It is crucial to note that the government will not necessarily choose this inventory level, however, since the political process will not generally attach equal weights to the wealth positions of all individuals.

In the light of the above analysis, it is apparent that Heller's point that foreign exchange reserves have a value to the community depends upon a particular assumption about the tastes and opportunities of the community; namely that the aggregate wealth curve is $\overline{W}$. In my paper I made the assumption, which I later qualified in a footnote, that the aggregate wealth curve was $\overline{F}$ and that an increase in the permanent stock of foreign exchange reserves will constitute an equal reduction in wealth. Had I had Heller's thought provoking comments when I wrote the paper, my preference would have been an assumption that the wealth curve is $\overline{W}''$, although analytical convenience would still have forced me to use my original assumption in the main body of the paper. Though some individuals benefit and others lose from a system of fixed exchange rates, I can find no convincing argument that such a system increases the production opportunities of the community and some quite convincing arguments that the contrary is true. Moreover, I have a difficult time visualizing a fixed exchange rate system as a consumption item.

I would agree with Heller that there is some optimum average or permanent inventory of foreign exchange reserves under a fixed exchange rate system though the optimum is second best if a flexible rate system would be more efficient.

The validity of Heller's point that a change in the stock of foreign exchange reserves is merely a reshuffling of assets depends critically on what kind of a deficit the public thinks is being financed. If the deficit is regarded as a transitory phenomenon, the permanent stock of foreign exchange reserves, as the public views it, will remain unchanged. Current reserves will change, as indeed they are supposed to, as a result of random variations in autonomous current and capital transactions, but the change will be transitory. The level of wealth will be unaffected. On the other hand, if, as I assumed in my paper, the community regards the disequilibrium as a "fundamental" one,

so that the permanent stock of foreign exchange reserves is changing, the level of wealth may be affected in either direction. Wealth will fall regardless of the direction of change of the permanent stock of reserves if the initial stock was a local aggregate wealth maximizing one such as F_1 or F_3. There is no particular reason to assume that the government will be holding this stock, however, since the political process need not weigh the wealth positions of all individuals equally. In this event, a change in the permanent stock of foreign exchange reserves may alter the aggregate wealth of the community in either direction.

Pareto Optimal Redistribution: Comment

By Paul A. Meyer and J. J. Shipley*

In discussing redistribution in the context of Pareto optimality, Harold Hochman and James Rodgers (hereafter H-R) have reopened a very interesting line of inquiry. However, the analytical simplifications which H-R felt required to make do much violence to problems they wish to study and imply results which, had they recognized them, they too would have considered absurd. In particular, their assumptions imply that the wealthy have an infinitely elastic demand to transfer income to the poor.

Mutt's transfer opportunity locus or budget line, which is repeated in their Figures 1–5 (H-R, pp. 546–50), is inconsistent with the situation which H-R set out to portray. In these graphs H-R have on the ordinate the income differential between the two individuals, Mutt and Jeff, and on the abscissa the amount of income which is transferred from the high-income individual Mutt to the low-income individual Jeff. The question is: given some initial income differential, $Y_M^0 - Y_J^0$, what is the shape of the budget line describing all the possible combinations of income differentials and income transfers, subject to the constraints that all the income given up by Mutt is received by Jeff and that Mutt's income never falls below Jeff's? Clearly, all the diagrams should show the budget line passing from $(0,\ Y_M^0 - Y_J^0)$ to $((Y_M^0 - Y_J^0)/2,\ 0)$, as in Figure 1. In this figure $OZ = Y_M^0 - Y_J^0$ is the initial differential and Y_T is the transfer to Jeff. The slope of the budget line is -2; a dollar transfer to Jeff reduces the differential by two dollars.[1] The maximum transfer occurs when the equality satisfies the constraint $Y_M \geq Y_J$.

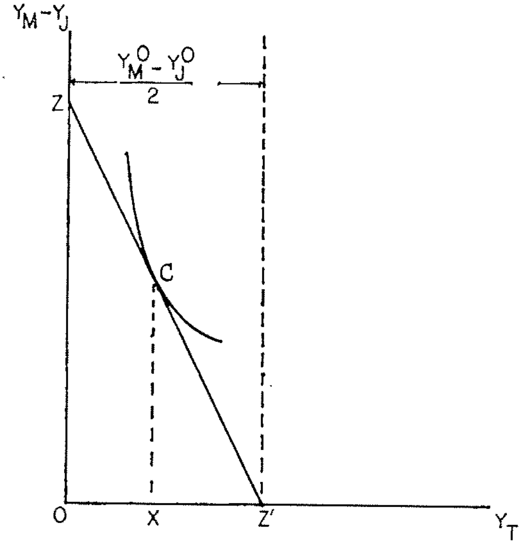


Figure 1

H-R draw a kinked budget line that becomes vertical after having an initial slope of -1. This error is minor, leading only to confusion. Given the correct budget line, as drawn in Figure 1, what point will Mutt pick? Though H-R state that Mutt's utility is a positive function of his and Jeff's incomes, they depict Mutt's indifference curve on this diagram in $(Y_T,\ Y_M - Y_J)$ space, and attribute Pareto optimality to point C. The ability to transform Mutt's utility surface into these dimensions from $(Y_J,\ Y_M)$ space follows from H-R's crucial assumption that the size of the transfer depends only upon the initial income differential. They fail to realize, however, that this assumption also restricts the shape of the indifference curves. Consequently, they commit a second error, this one much more serious. In fact, their results regarding the possible values of E_M, Mutt's transfer elasticity, are all inconsistent with their assumptions.

To show this, suppose contrary to H-R that Mutt's utility surface, $U_M = f_M(Y_J, Y_M)$, is mapped in $(Y_J,\ Y_M)$ space, as in Figure 2.

* The authors are, respectively, associate professor of economics and lecturer in economics at the University of Maryland.

[1] Algebraically, $b = (Y_M - Y_J - (Y_M^0 - Y_J^0))/Y_T$, where b is the slope of the budget line and Y_T is income transferred. By definition, $Y_T = Y_M^0 - Y_M = Y_J - Y_J^0$. Substitution of the definitional equations in the first equation shows that $b = -2$.

Pick some point A, above the 45° line, and let this be (Y_J^o, Y_M^o). The vertical distance between A and the 45° line is the initial income differential. The transfer possibility curve in this diagram, equivalent to line ZZ' in Figure 1, is the line with a slope of -1, passing from A to the 45° line. While a dollar transfer to Jeff reduces the differential by two dollars, the increase in Jeff's income equals the decrease in Mutt's income. H-R tell us that Mutt can increase his utility by transferring income to Jeff. Mutt maximizes his utility by moving from A to B, i.e., by reducing the income differential from $Y_M^o - Y_J^o$ to $Y_M^E - Y_J^E$ through the transfer of income in the amount of $Y_M^o - Y_M^E$ from Mutt to Jeff. H-R now ask how Mutt's preferred income transfer would change if the initial income differential had been bigger, say at point C. Again the transfer possibility curve is the line with slope of -1 passing from point C. And again Mutt can increase his utility by transferring some of his income to Jeff. Where will the new Pareto optimum be?

Before attempting to answer that question, consider point A'. By construction, A' lies on a line parallel to the 45° line and, hence, represents the same income differential as point A. From H-R's assumption that the transfer is a function of only the initial differential, Mutt will prefer to transfer income to Jeff and will end up at point B', the initial and equilibrium income differentials and the income transfer being the same as between points A and B. Line BB' also is parallel to the 45° line. Now consider point C. If Mutt's indifference curves are shaped in the usual fashion, and nothing suggests otherwise, Mutt will also prefer to move from point C to point B'. By extension, from any point to the left of the line BB' Mutt will always prefer to transfer back to a point on that line. From points to the right of BB' no transfer from Mutt to Jeff can raise Mutt's utility, and none will occur.

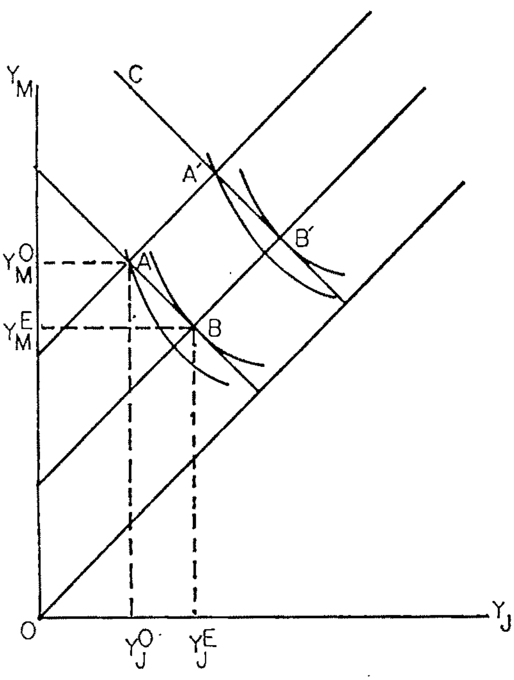


FIGURE 2

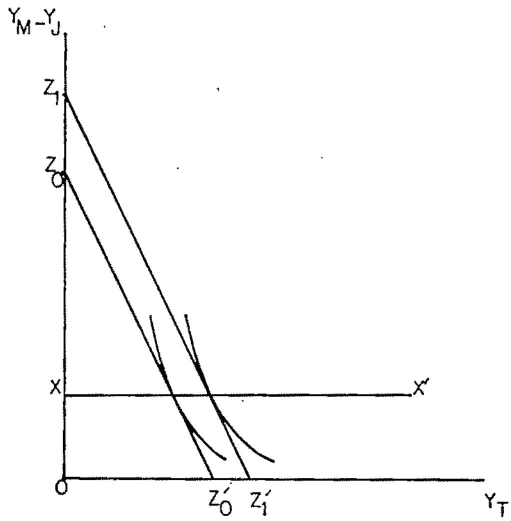


FIGURE 3

This implies that in $(Y_T, Y_M - Y_J)$ space the slopes of Mutt's indifference curves are constant along a horizontal line. The IC curves, which are the loci of equilibrium differential-transfer points drawn on H-R's Figures 2–5, are not consistent with this implication. Let the initial income differential vary, as illustrated by Z_oZ_o' and Z_1Z_1' in Figure 3. Pareto optimality, if we are to follow H-R's restrictions, always requires a transfer big enough to return the income differential to $Y_M^E - Y_J^E$. The IC curve

OXX' is the only one possible. By definition, Mutt has an infinitely elastic "demand for increments to Jeff's income."

All this should be sufficient to suggest putting aside H-R's simplifications and calculations and beginning to consider the problem of optimal redistribution in a setting which is less absurd. Unfortunately that is not within the scope of a comment.

REFERENCE

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–57.

Pareto Optimal Redistribution: Comment

By RICHARD A. MUSGRAVE*

The analysis by Harold Hochman and James D. Rodgers of redistribution qua exercise in Pareto optimality offers a helpful and down-to-earth addition to other recent efforts in this direction.[1] Now that the case has been made, it is indeed difficult to see why this aspect had so long been neglected. However, the following comments seem in order.

I

The degree of "redistribution" which occurs in the context of the Hochman-Rodgers scheme is a function of a) people's rate of substitution between the satisfaction derived from retaining income and that derived from giving it, and b) the "initial" distribution of earnings which exists before giving occurs. Whatever the values of a), the outcome will differ depending on b). Such, at least, will be the case unless everybody's rate of substitution is such that complete equality results. Pareto optimal redistribution thus constitutes a *secondary redistribution* which depends on the initial distribution of earnings. This distribution is determined by such factors as inheritance, earning capacities, education, and market structure. It may itself be changed through the political process. Such changes, referred to here as *primary redistribution*, are not a matter of voluntary giving, but of taking.

How is this primary redistribution decided upon? To the extent that it operates within the legal framework, it is performed through the voting mechanism. At a normative level, this explanation is not very helpful, however, since it merely raises the next issue, i.e., how the distribution of votes and the voting rules are to be decided on. Eventually, the problem becomes one of social-contract determination. At a positive level, primary redistribution depends on the social structure and balance of power between income groups. As these change, corresponding changes occur in the voting decisions and/or voting rules.

II

Turning to an application of these concepts, how can the primary or non-consent component be distinguished from the secondary or Pareto component? Private redistribution or charity in the United States now accounts for about $12 billion.[2] Assuming (somewhat unrealistically) that this giving is truly voluntary, it must be interpreted as being of the secondary type. But how should one interpret the much larger block of budgetary redistribution, which may well exceed $30 billion?[3]

* Professor of economics, Harvard University.

[1] See the Hochman and Rodgers *Review* article and the references there given. To these should be added S. C. Kolm's paper on the optimal production of justice.

[2] Charitable contributions claimed for itemized deductions in 1969 amounted to about $11.5 billion and the resulting tax revenue loss to over $3 billion, leaving about $8 billion of privately financed contributions. To this let us add, say, $3.5 billion to account for contributions made by returns using the standard deduction, giving a total of about $12 billion of private contributions. This figure does not cover interfamily giving. Using the family as a unit, this is not considered a part of redistribution.

The figure of $3.5 billion for nonitemized deductions is based on the following conjecture. Applying the ratio of contribution to *AGI* for returns with itemized deductions to returns with standard deductions, we obtain an upper limit of $6.5 billion. This, however, is surely an overestimate since nonitemizers may be expected to have a lower contribution ratio. The figure of $3.5 billion sets this ratio at roughly 50 percent of that for itemizers.

[3] This amount is the estimated total gain in net benefits (gain from public expenditures minus loss from taxation) on the part of those whose net benefits are positive. Obviously such estimates are difficult to obtain since they imply the entire scope of both tax and expenditure incidence, including the imputation of benefits from public services. For a discussion of the problems involved in estimating expenditure benefits, see I. Gillespie. The figure of $30 billion corresponds to Gillespie's estimate of $17.5 billion for a 1960 budget total of $133.6 billion, raised to allow for a budget total of $300 billion. The underlying figure of $17.5 billion

Hochman and Rodgers assign this entire amount as well to the secondary category. This follows from their assumption of "... an institutional setting in which free-riding, i.e., strategic behavior, is precluded so that the political mechanism through which interdependence is internalized accurately reflects the distributional preferences of individuals in this regard" (p. 548). This permits them to treat the large number case as if it involved small numbers, but the assumption is not permissible.

Rather, the situation is similar to that of provision for social goods. Let us specify that Mutt's satisfaction from giving to Jeff derives from the fact that Jeff's consumption is increased. Mutt, therefore, derives the same satisfaction if the transfer originates with Sam. With small numbers, Mutt and Sam get together and negotiate their giving to Jeff.[4] But the situation differs if Mutt is confronted with 60 million Jeffs and Sams (households) instead of one. He will readily conclude that his giving will not contribute significantly to the welfare of the Jeffs, so that uncooperative behavior on his part will not meet with retribution by the Sams. Some way must be found by which preferences are revealed and concerted action is agreed upon. Hence secondary (Pareto optimal) redistribution must be implemented through the political process, just as such a process is needed for the provision of social goods. This is why secondary redistribution extends into the budgetary process.

But if so, how can it be distinguished from primary redistribution which also operates through the budget? The distinction is subtle but nonetheless important. In its secondary redistribution component, the budgetary process reflects an attempt to approximate what in the small number case would be accomplished voluntarily by individual bargains, and without compulsion. Since numbers are large, compulsion (the mandatory application of the voting decision) is needed to secure revelation of preferences, but it is a necessary evil only, not an objective in itself. Budgetary redistribution in its primary component, on the contrary, involves compulsion by its very nature: the Jeffs succeed in taking from the Mutts and Sams against their will.

The existing pattern of budgetary redistribution, therefore, includes both a secondary and a primary component. While I suspect that a substantial part is of the primary type, there is no simple test by which the two components may be distinguished in practice. Indeed, it would not be realistic to think of particular redistribution dollars as being assignable entirely to one or the other group. Both aspects are present and intertwined in determining the redistributional pattern.

III

Finally, a word about the bearing of secondary redistribution on my distinction between allocation and distribution policy, a distinction which was formulated with primary redistribution in mind.[5] Allowing for the possibility of secondary redistribution moves part of the distribution issue into the framework of Pareto optimality, and in this sense aligns it with the "allocation branch." But this does not mean that policies to secure an efficient *product* mix must involve distributional considerations, and vice versa.

The Hochman-Rodgers approach looks at Pareto optimal redistribution and I think correctly so, in terms of money income. Thus, two Pareto games are being played in two adjoining boxes: One involves the distribution of money income, while the other involves the determination of the product mix with any given distribution. Mutt derives welfare from Jeff's increase in income, but accepts Jeff's preference pattern and

assumes allocation of benefits from general expenditures (expenditures which do not permit specific allocation) in line with income. If such benefits are allocated on a per capita basis, the level of redistribution is nearly doubled.

[4] The situation is similar if Mutt derives value from Jeff's giving because this maintains his net (after-giving) income position relative to Jeff's. While Sam's position is improved, this does not carry adverse implications because he is too far down the line. If, on the other hand, Mutt's satisfaction derives from the fact that the gift originates with *himself*, he will be indifferent to gifts by others. In this case secondary giving will not call for budgetary action and be wholly private.

[5] See my *The Theory of Public Finance.*

decision on how to divide this income between oranges and apples.[6] As long as the interpersonal utility argument in Mutt's utility function is individualistic, my essential point of separation between allocation and distribution policies is retained. Redistribution should be handled in terms of income transfers, and not by pricing policy. I shall be pleased henceforth to divide my Distribution Branch into two compartments, D' being Pareto optimal and D'' primary, but this is all that is needed.

The situation differs if Mutt values Jeff's consumption of oranges and apples in accordance with his own (Mutt's) rates of substitution. In this case, allocation and distribution are inseparable. The provision for merit goods may be interpreted in these terms, and is frequently linked with redistribution.[7] Equivalent in nature to a cash grant earmarked by the donor for specified uses, redistribution in kind may be of the secondary or primary type, but there is some presumption that it is secondary, since the claimant would hardly wish to restrict his choice in the primary case.

Redistribution, it appears, involves a complex set of transfers, primary or secondary, in cash or in kind, and operating both inside and outside of the budget. While the exploration of cash-giving in terms of Pareto optimality is instructive, it covers but a part of the problem and cannot claim to explain the entire phenomenon at hand.

[6] Dr. Mutt, at his most sophisticated, realizes that his welfare gain from giving depends on Jeff's welfare from receiving, and that the latter depends not only on changes in money income but also in relative prices. In this sense the allocation effects of giving are linked to the distribution effects. Dr. Mutt will take into account prevailing price and market structures. In this connection, the problem is the same as that discussed in my defense of separation between allocation and primary redistribution issues. See my 1966 article.

[7] I have suggested this approach as an explanation to the prevalence of merit goods in my 1966 article.

REFERENCES

I. Gillespie, "Effect of Public Expenditures on the Distribution of Income," in R. A. Musgrave, ed., *Essays in Fiscal Federalism*, Washington 1965.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–57.

S. C. Kolm, "The Optimal Production of Justice," in J. Margolis and H. Guitton, eds., *Conference on The Analysis of the Public Sector, Barritz, 1966*, New York 1969.

R. A. Musgrave, *The Theory of Public Finance*, New York 1958.

———, "Provision for Social Goods," in J. Margolis and H. Guitton, eds., *Conference on The Analysis of the Public Sector, Biarritz, 1966*, New York 1969.

Pareto Optimal Redistribution: Comment

By Robert S. Goldfarb*

In their recent article in this *Review*, Harold Hochman and James D. Rodgers use the assumption of interdependent utility functions to build a theory of Pareto optimal redistribution. This note presents a simple specification of their analysis which clarifies some aspects of their procedures and yields additional results.

Hochman and Rodgers assume that individual X derives additional utility not only from increases in his own income, but also from increases in individual Z's income, i.e., that $U_x = f(Y_x, Y_z)$. As they point out, in such a world there may exist income transfers which make everyone better off. Income can be taken away from X and given to Z, and both X and Z may be better off. That is, some income transfers are Pareto optimal; they make at least one person better off, while hurting no one.

For illustrative purposes, Hochman and Rodgers discuss a two-person case involving Mutt and Jeff (our X and Z), and then jump to the N-person case without much discussion of the various possible forms of utility functions in the N-person case. We will suggest one such form which we believe is intuitively reasonable. This form generates the result that private charity will not bring about all Pareto optimal redistributions; the government, through its powers of coercion, can tax and spend to bring about income transfers that leave everyone better off.

This result provides a more explicit basis for some of Hochman's and Rodgers' procedures. They argue that the government may want to use the tax mechanism to bring about Pareto optimal redistribution, but do not demonstrate explicitly that these redistributions will not necessarily be made completely by private charity. We are able to show this by specifying in a reasonable way the forms of utility functions in our N-person world.[1]

Assume a society of N individuals, $K < N$ of whom are poor. Assume further that the $N-K$ "rich" individuals have utility functions which value gains in income to themselves and to the poor.[2] Thus,

$$U_{R_i} = f(Y_{R_i}, Y_1, Y_2, \ldots, Y_K) \tag{1}$$

where R_i represents rich individual i and $Y_1 \ldots Y_k$ are the incomes of each of the poor. Note that this specification of the utility function of our "typical" rich individual implies that he is indifferent to income losses (or gains) of other rich individuals; these gains or losses are not reflected in his utility function. Our rich man cares about his own welfare and the welfare of the poor, but not about the welfare of other rich people.

Suppose that the government proposes a \$1 tax on all rich individuals, the proceeds to be distributed equally among all of the poor individuals. It can be shown that cases arise where our typical rich individual, R_i, would be better off by the tax, even though he would not choose to give (i.e., would not be made better off if he gave) \$1 to the poor on his own. If the government levies the \$1 tax, $(N-K)$ dollars will be collected, so that the

* The author is assistant professor of economics at Yale University. He thanks Marsha Geier Goldfarb, George A. Hay, and John McGowan for helpful comments. Remaining errors are his own.

[1] Hochman and Rodgers refer to this issue in fn. 4, stating that "... In the N-person case, however, individuals, unless coerced, may choose to be 'free riders' and it is the incentive to behave in this way that may be viewed as the raison d'etre of government.... And when we turn to the N-person case, we assume that the possibility of voluntary redistribution through private charity has been exhausted, thus focusing on the incremental redistributive activities carried out under public auspices" (p. 543). Our approach in this comment is to show that, once voluntary redistribution is exhausted, the incremental redistribution activities left over can in fact be Pareto optimal. Put differently, a somewhat peculiar variant of the free rider problem is likely to exist, justifying government transfer activity.

[2] Being rich may be defined as having an income greater than $\bar{Y}$; being poor may be defined as having an income less than Y^*, where Y^* is less than $\bar{Y}$.

income of each poor person rises by $(N-K)/K$ or $N/K-1$. R_i's utility gain from this is

$$\left(\frac{N}{K}-1\right)\sum_{j=1}^{k}\frac{\partial U_{R_i}}{\partial Y_j}$$

His utility loss from the additional \$1 in taxes is

$$\frac{\partial U_{R_i}}{\partial Y_{R_i}}$$

so the net gain (or loss) from the tax and redistribution is approximately

$$(2)\qquad \left(\frac{N}{K}-1\right)\sum_{j=1}^{k}\frac{\partial U_{R_i}}{\partial Y_j}-\frac{\partial U_{R_i}}{\partial Y_{R_i}}$$

If, instead, R_i had given through charity (and the \$1 had been distributed in the same way), the gain (or loss) would have been approximately

$$(3)\qquad \frac{1}{K}\sum_{j=1}^{k}\frac{\partial U_{R_i}}{\partial Y_j}-\frac{\partial U_{R_i}}{\partial Y_{R_i}}$$

Clearly, (1) is larger than (2) so long as N is $>(K+1)$(i.e., so long as there is more than one rich person). If all rich people have this typical utility function, Pareto optimal redistribution may be possible through taxation even when there is no voluntary Pareto optimal redistribution open to an individual rich person.[3]

Let us put the argument in common sense terms. If Seymour Rich gives \$1 to charity, he knows that K poor people will each receive \$$1/K$. This gives him some satisfaction, but the loss of the \$1 grieves him. If, on the other hand, the government comes along and taxes him (and every other rich man) \$1 his grief is the same as in the charity case, but his satisfaction is much larger; he knows that, with his payment of \$1 in taxes, the poor each receive \$$(N-K)/K$, a number much larger than the \$$1/K$ they get in the charity case. Thus, he might be unwilling to donate the \$1 to charity (the satisfaction being less than the grief), but very willing to vote for the \$1 tax (the satisfaction is bigger than in the charity case, and therefore can be greater than the grief).

The reader should note that our reasoning does not depend on the way the charity and the government split the money up among the poor. The same argument would hold, for example, if the rule which both the government and the charity followed was to give all the collected money to poor individual F. Under the charity scheme, F would get \$1 as a result of Rich's \$1 contribution, whereas under the tax scheme, he would get \$$(N-K)$ at the cost of only \$1 to Rich. Thus, Rich may favor the tax over the charity and therefore may vote for (want) the tax even though he does not make the charity contribution. The only thing our argument depends on is that the government and the charity follow the same rule for distributing to the poor; if the government gives all the money to F, the private charity must do so also.

The reason for the possibility of gains through collective action is clear. When individual R_i (or Seymour Rich) pays his \$1 tax, he knows that other rich individuals must also pay, and that the total collected will be given to the poor. Thus, in paying his \$1, he is assured of a much larger transfer going to the poor. Since his utility is a function of the number of dollars the poor gain, but is not a function of the number of dollars other rich people lose, he is better off if all other rich people pay (i.e., the government scheme) than if they do not (i.e, charity). In short, if the utility functions of the rich have the particular form we postulated, a "large" government poverty program may be Pareto optimal even though private charity programs for the poor do not receive any contributions.[4]

Notice that our result about the desirability of government action could be altered by two quite different qualifications. First, if there exists a mechanism by which Seymour Rich can sign a "charity contract" which says "I will contribute \$1 if all $N-K$

[3] This kind of argument is used in a different context in Stephen Marglin's article.

[4] Notice the public good nature of each redistribution. It benefits not only the giver and the receiver, but also all rich people with the specified utility functions. It is this "public goodness" of redistribution which accounts for the Pareto optimality of the tax.

of the other richniks do," then all Pareto optimal transfers will be made by the private sector; Pareto optimality is no longer an argument for government-imposed redistributions. Second, suppose that Seymour Rich's utility function includes the income and welfare of his rich friends as well as that of the poor. Since a tax on the rich hurts his friends, and this "hurt" enters his utility function, the dominance of taxes over charity is no longer clear-cut. It will all depend on how strongly income losses to his friends count against income gains to the poor. Oddly enough, adding more interdependencies (i.e., putting the incomes of additional people in Seymour Rich's utility function) may weaken the case for government action.

Let us return to our original world, ignoring the two qualifications just mentioned, and see whether any subsidiary results flow from the analysis. One additional result concerns the population characteristics of societies more likely to have available a governmental Pareto optimal transfer scheme. We have argued that it is possible, given the specified types of utility functions. that situations arise in which the government may be able to make everyone better off by redistributing income. Can we say anything about what type of society is most likely to give rise to these situations? Suppose we compare two societies with equal numbers of poor people; then the larger the number of rich people (other things equal),[5] the larger the satisfaction to an individual rich person from the same tax scheme. That is, the more rich people there are, the larger will be the transfer to the poor which results from a $1 tax on each rich person. Thus, in a society with N poor people but "few" rich people (i.e., $N-K$ is "small"), Seymour Rich will not view taxes as having a large advantage over charity, whereas in a society with N poor people but "many" rich ($N-K$ "large"), there would be a bigger advantage to taxes over charity. A tax scheme which Seymour favors is more likely to be available in the second society than in the first. For cases with the same number of poor people, the smaller the percentage in poverty, the more favorably a poverty program would be received. The less pressing the poverty problem is in relative terms, the more likely it is to be enacted.[6]

A second subsidiary result concerns tax schemes. The public finance literature contains numerous attempts to justify a progressive tax rate. Does a system of utility functions such as the one described above justify progressive taxation? Any switch from a tax system which taxes the poor to one which taxes the rich is a move toward a Pareto optimum under the above assumptions about utility functions so long as

$$\left(\frac{N}{K}-1\right)\sum_{j=1}^{k}\frac{\partial U_{R_i}}{\partial Y_j} > \frac{\partial U_{R_i}}{\partial Y_{R_i}} \tag{4}$$

That is, a shift in the tax burden is Pareto optimal so long as a $1 income loss to R_i does not cause more disutility to R_i than the satisfaction he gets from easing the tax burden on the poor by $N-K$ dollars. Since progression is a way of shifting the burden of a required tax yield from the poor to the rich, a move towards progression could be (but need not always be) a Pareto optimal movement if the postulated type of utility func tions exists.

REFERENCES

H. M. Hochman and James D. Rodgers "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–56.

S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quart. J. Econ.*, Feb. 1963, *77*, 95–112.

[5] "Other things equal" would imply that all additional rich individuals have the same utility functions as the first set of rich individuals, that per capita income of the rich (and the poor) in the larger society was the same as the income of the rich and the poor in the smaller society, and so forth.

[6] Many writers have remarked on the somewhat startling fact that the United States only became really concerned about poverty in the 1960's, when there was much less of it than ever before. The theory presented here suggests one type of explanation for this phenomenon. Far more appealing sociological explanations of this previous "lack of concern" are available however. These sociological theories sometimes imply that utility functions have shifted recently towards a greater concern for the poor.

Pareto Optimal Redistribution: Reply

By Harold M. Hochman and James D. Rodgers*

The comments raise two objections to our discussion of "Pareto Optimal Redistribution." The first objection, discussed by Paul A. Meyer and J. J. Shipley, relates to our two-person model, while the second, introduced by Richard A. Musgrave, deals with the extent to which redistribution can be explained in terms of utility interdependence and the analytic device through which we extend our analysis from the two-person to the N-person case. Robert Goldfarb's comment contains a partial reformulation of our N-person model, which establishes the link between Stephen Marglin's analysis and ours, thus indicating how some of the difficulties that Musgrave cites can be resolved.

I. *Meyer and Shipley*

The Meyer and Shipley comment, though helpful in that it points the way to a needed clarification of our model, is overstated. Their first point, which alleges that the slope of the budget line in our Figure 1 should be -2 rather than -1, betrays a failure to understand what the vertical axis of our diagram measures. This axis measures the *initial* income differential ($(Y_M-Y_J)^0$, the intercept of the budget line), and *the portion of the initial income differential which Mutt retains for own use*. This, of course, is $(Y_M-Y_J)^0-Y^T$, where Y^T is Mutt's transfer to Jeff. *The vertical axis does not measure the post-transfer income differential.* The increase in Jeff's income (above its initial level) which the transfer brings about is measured on the horizontal axis. Hence, the "slope of the ZZ' segment of the budget line is -1, since a given size transfer to Jeff reduces the amount of income that Mutt retains for his own use by the same amount" (Hochman and Rodgers, p. 546). This should be clear from our discussion of what is measured in Figure 1, as well as from our definition of Mutt's equilibrium at E, "where the marginal utility of a dollar of own-consumption equals the marginal utility (to Mutt) of a one dollar increment in Jeff's income." (p. 546).[1] Unfortunately, however, our labeling of the vertical axis (as indicating "Y_M-Y_J") was ambiguous and an understandable source of confusion.

The second point in the Meyer and Shipley comment is that our Figures 2–5, illustrating the alternative configurations of Mutt's preferences, are technically incorrect, because our "*IC* lines which are the loci of equilibrium differential-transfer points . . . , are not consistent with (our) assumptions." Our assumptions, they argue, imply that "Mutt has an infinitely elastic demand for increments to Jeff's income."

The first half of this criticism is technically correct, though it imputes more importance to the formal characteristics of our model than we intended and is easily repaired without altering the sense of the argument. The inconsistency that Meyer and Shipley identify derives from our *ceteris paribus* assumption that the size of the transfer which Mutt wishes to make to Jeff depends only on the initial differential between their money incomes. Specified in general terms, Mutt's utility function, in our model is $U_M=f_M(Y_M, Y_J)$. Our analysis, however, focuses

* Director of studies in urban public finance, The Urban Institute, and department of economics, Pennsylvania State University, respectively. The authors wish to acknowledge the support of the National Science Foundation, under Research Grant GS-2805 to the Urban Institute. They are indebted, for useful insights, to James M. Buchanan, Francesco Forte, William Hamburger, William McEachern, and A. Mitchell Polinsky.

[1] We note, in passing, one way of improving our representation. In place of our restriction precluding transfers which reverse the distributional ordering, we might make the reasonable assumption that the slopes of Mutt's indifference curves are zero or less where they intersect a 45° ray from the origin. By doing this, we eliminate the need for the $Z'Z''$ segment of our budget line and the strict prohibition of some Pareto optimal transfers that our original formulation implies.

on the subset of (Y_J, Y_M) space with origin $Y_M = Y_J$ and coordinates which measure the portion of the initial differential which Mutt retains for own use and Y_E^T, Mutt's equilibrium transfer to Jeff. In the two-person case, the assumption that Mutt's transfer depends only on the differential restricts Mutt's utility function to a particular form, which implies, as Meyer and Shipley point out, that Mutt's IC line must have a constant positive slope with an intercept on the positive portion of the vertical axis. This holds true for a diagram representing all of (Y_J, Y_M) space, such as Figure 2 in the Meyer and Shipley comment, and for the diagrams which our paper employs in illustrating the two-person model.[2] It must be recognized, however, that the two-person case, in our view, is just a building block on which the N-person extension can be based. Provided that our diagrams are viewed in an N-person context and that it is recognized that Mutt's desired transfer to any given Jeff at a particular point in time is not independent of the incomes which other Jeffs in the community receive, the Meyer-Shipley criticism does not invalidate the diagrams depicted in our Figures 2–5.[3]

While our paper did not explicitly recognize how severe a restriction our formulation of the two-person case imposes on Mutt's preferences, relaxation of this restriction leaves our essential argument intact. Moreover, as we noted, our formulation of the two-person case can be repaired in either of two ways. First, we might simply admit that the utility function differs in form for each of the diagrams and for each of the alternative numerical examples; conceptually, however, this is not very satisfying. As an alternative, we can redefine the conceptual experiment through which the amount of the Pareto optimal transfer is derived, so that the restrictive IC outcome that Meyer and Shipley describe does not result, thus eliminating the problem. Originally, in setting up our model, we did not distinguish between changes in $Y_M - Y_J$ (changes in the intercept of the budget line on the ordinate) that result from (1) changes in Y_M and (2) changes in Y_J, implying thereby that our analysis was capable of registering the effects of diagonal movements in Y_J, Y_M space.[4] This is the source of the restrictive outcome. By redefining our conceptual experiment, so that the changes in $(Y_M - Y_J)^O$ which we consider derive only from changes in Y_M, with Y_J held constant (or the reverse,) diagonal movements in (Y_J, Y_M) space are ruled out as objects of our analysis and our two-person case becomes fully consistent with all of the IC lines that our paper goes on to discuss.

So much for our discussion of this problem. We might comment briefly, however, on the Meyer and Shipley claim that our assumptions imply that Mutt has an "infinitely elastic demand for increments to Jeff's income" (p. 990). Presuming that their reference, in saying this, is to income elasticity (since the price of a transfer, i.e., the slope of ZZ', is always -1 in the two-person case), this statement cannot be correct, as a matter of simple arithmetic. There is no way in which the response of Y^T to a change in Mutt's income (or, for that matter, to a change in $(Y_M - Y_J)^O$) can be infinitely large.

II. *Musgrave*

The essential argument of our paper was that Paretian welfare economics, contrary to the conventional view, can help to resolve dissatisfaction with the size distribution of income. Once the interdependence of utility functions is admitted, the range of distribu-

[2] Note that the representation in Figure 3 of the Meyer and Shipley comment is incorrect. Their budget line should have a slope of -1 and the IC line should have a positive slope.

[3] Suppose, for example, that Mutt has an income of \$5,000 and that two Jeffs (J_1 and J_2) have incomes of \$2,000 and zero, respectively. Let the size of Mutt's preferred transfer to J_1 be $Y_{J_1}^T$ and the size of his preferred transfer to J_2 be $Y_{J_2}^T$. Now suppose J_2's income rises to \$2,000 while J_1's remains at \$2,000. As a result the Pareto optimal transfers become $Y_{J_1}^{T*}$ and $Y_{J_2}^{T*}$. It is not generally true that $Y_{J_1}^{T*}$ need equal $Y_{J_1}^T$ nor that $Y_{J_1}^{T*}$ need equal $Y_{J_2}^T$. Hence, in an N-person model our figures are not inconsistent with our assumption that transfers depend on the initial income differential.

[4] We had claimed, in fn. 12 on p. 549, that it makes no difference whether a change in the initial income differential derives from a change in Y_M or a change in Y_J.

tional conflict is narrower than economists have supposed, to the extent that interdependence implies that the initial or market-determined income distribution is not coincident with the distribution of welfare. The relevant implication of interdependence is that some portion or portions of the utility-possibility locus, in the two-person case, may slope upward, so that some of its points (each of which corresponds to an initial distribution of money income) are dominated.

Musgrave argues that this line of reasoning, while correct, resolves only a part of the social issue of distributional justice. He reclassifies redistribution under two headings: "primary" redistribution, involving a choice among points on the efficient segment of the utility-possibility frontier; and Pareto optimal redistribution (which Musgrave calls "secondary") involving a movement from an inefficient (positively-sloped) to an efficient portion of this locus.

Primary redistribution, at the normative level, is a matter of "social contract," which evolves from the process of constitutional choice through which a society establishes its structure of common and individual rights to human, physical, and natural resources. Any society, in doing this, must settle not only on an optimal distribution of current income, but, more generally, on rules for distributing rights and opportunities to earn and expend money income and consume leisure. The "distributional" outcome, in a total sense, is achieved through a bargaining process, in which "free" individuals, seeking agreement, revise their preferences and/or opt in or out of the political community. The society which is produced by the contracting process, emerging from a consensus of free individuals, has institutional (including "distributional") characteristics that are, by definition, Pareto optimal (in the sense that those persons who cannot subscribe to its constitution choose exclusion from the community). Interpreted as an element in this process, "primary redistribution" cannot be a matter of "taking" rather than "voluntary giving," as Musgrave claims (p. 991). Taking, per se, is in fact inimical to the notion of the social contract that characterizes the constitutional choice process (which is in turn the basis of modern theories of democracy).

The issues that primary redistribution involves are therefore broader than the traditional public finance literature admits, or than Musgrave's dichotomy implies. Still, if the only distributional issue addressed is the apportionment of current income, as in the conventional framework, Musgrave's distinction between primary redistribution, which makes some individuals worse off, and secondary redistribution, which makes no one worse off, is well taken. But the terms primary and secondary, which Musgrave has chosen to designate such distributional adjustments, are misleading, for the logical ordering of the two components of the redistributive process that they imply is incorrect. Specifically, one cannot determine in welfare terms, whether the initial or market-determined distribution of income is satisfactory *until* Pareto optimal transfers have occurred.

However, when we turn to the very different question of real world redistribution, we can certainly agree with Musgrave that its primary and secondary components are "intertwined in determining the redistributional pattern" (Musgrave, p. 992). To argue that all redistribution is in the Pareto optimal category would be unrealistic, just as it would be unrealistic to argue that all redistribution is primary or that all charitable contributions are voluntary. Surely, we did not intend such an impression, and Musgrave, in criticizing our assignment of all government redistribution to the Pareto optimal category, has misinterpreted us. The first and second sections of our paper dealt with a hypothetical situation in which only *Pareto optimal* transfers were considered. Then, for argument's sake, its third section asked what pattern of utility interdependence might be implied if (again hypothetically) the actual pattern of fiscal residuals were Pareto optimal. But it did all this in an explicitly heuristic spirit, without false pretense to realism.

In criticizing our discussion of the N-person case, Musgrave contends that treating "... the large number case as if it

involved small numbers . . . is not permissible" (Musgrave, p. 992). However, whether such a simplification is admissible or not depends on one's assumptions about the precise character of the motivation that underlies the income transfers. One possible case, as Musgrave suggests, is that in which Mutt's satisfaction derives "from the fact that the gift originates with himself." But, Mutt's benefit, alternatively, might derive from the improvement in the recipient's welfare which the transfer produces. In this latter case Mutt's motivation might be genuine concern for the recipient's plight *or* the belief that the gift will reduce the likelihood that the recipient will participate in antisocial behavior (e.g., in riots which affect the value of Mutt's private property). If each benefactor's benefits derive only from transfers that he makes himself and not from transfers which other persons make, simple aggregation of pairwise transfers is the proper method of analysis in the N-person case and our straightforward extension of the two-person model to the case of a community-at-large is appropriate. Since the market would not "fail" in this case, as Musgrave notes, a Pareto optimal outcome is attained through private redistribution, and the intermediation of the fiscal structure is not required. In the second case, however, a Mutt's transfer to any particular Jeff yields spillover benefits to other Mutts, and, unless highly qualified, the simple aggregation of pairwise transfers is not a correct procedure, except as an explicit matter of heuristic convenience. Transfers are a collective good which benefit all individuals in Mutt's income status vis-à-vis the recipient, implying an argument for coerced redistribution that is strictly analogous to the argument that "national defense" must be financed through the fiscal process. When there is a large number of potential benefactors, the attainment of a Pareto optimum through voluntary action is difficult, inasmuch as each participant has a private incentive to remain a "free rider" and stands to gain from a collective agreement which coerces him into participating in the transfer process. In these circumstances, attainment of a Pareto optimum requires that all marginally affected Mutts must share in the transfer to a particular Jeff, with the cost of increasing a given Jeff's income by one dollar being apportioned among the Mutts (assuming application of the benefit principle) on the basis of their marginal evaluations.[5]

Regarding the implications of our paper for Musgrave's division of the public sector into allocation and distribution branches, several comments are in order.

Where the interrelationships among utility functions are such that a large number of non-participants benefit from the process of redistribution, interpersonal transfers pose exactly the same kind of "market failure" problem as "national defense." Such transfers are, therefore, a collective good, albeit one that is not resource-using. Thus, in Musgrave's schema, they should logically be provided for by the allocation branch, since, concisely stated, its function is to carry out the public activities needed to assure a Pareto optimal allocation of resources.

At least for classification purposes, we see no difference between the case in which Mutt's benefit derives from an increase in Jeff's income or welfare and the case in which Mutt's benefit derives from an increase or decrease in Jeff's consumption of a particular commodity. Contrary to Musgrave's claim, the issue of whether a money or in-kind transfer is appropriate does not, in itself, indicate whether allocation and distribution are separable or inseparable. If each of the interested Mutts permits himself to be coerced into financing an in-kind transfer to Jeff (in the form of a consumption subsidy) that induces him to alter his consumption pattern and Jeff accepts the transfer on their terms, we have evidence that all of the individuals involved, including Jeff, have moved

[5] We reemphasize, however, that the issue of the appropriate formulation of the N-person case is separate and distinct from the main point that we wished to demonstrate with our simplifying assumptions, that is, the normative distributional implications of utility interdependence. At the same time it should certainly be clear that a realistic treatment of the N-person case, as suggested by Musgrave, and developed in part by Goldfarb, is a natural extension of our analysis and adds to its relevance.

to preferred situations. Jeff's acceptance, of course, unquestionably implies that he is better off than he would have been with *no transfer at all*. It is this *ex ante* state of affairs, rather than what the welfare outcome would have been if the transfer had been in general purchasing power, that is the appropriate basis of comparison. Transfers of general purchasing power, in Musgrave's terminology, can be *either primary* (in which case they are purely a distributional phenomenon) or *secondary* (in which case their allocative and distributive aspects are inseparable). However, one could argue that *all* earmarked or in-kind transfers are of the secondary type since recipients, if they had the power to institute redistribution in their favor, would prefer cash. An estimated 43 percent of all federal aid to the poor in 1968 took the in-kind form (see M. S. March).[6]

III. *Goldfarb*

While Goldfarb's useful insight into the nature of the N-person case requires no rejoinder,[7] some aspects of his comment, on which we shall dwell briefly, trouble us.

Goldfarb remarks that "In short, if the utility functions of the rich have the particular form we postulated, a 'large' government poverty program may be justified on Pareto optimization grounds, even though private charity programs, for the poor do not receive any contributions," (p. 995). This suggests a failure to understand why some collective wants are satisfied voluntarily, through charity or other private means of provision, while others must be provided for through the fiscal process. The size of a redistribution program in itself indicates nothing at all about whether it is justifiable in terms of the Pareto criterion, regardless of whether the *mechanism* through which it is put into effect is private or public. What matters is whether the collective marginal evaluations warrant the program. Whether private or public means of provision are appropriate depends on the costs of obtaining voluntary agreement. The real thrust of Goldfarb's observation, then, is that the free-rider problem will make the level of voluntary transfers suboptimal.

Moreover, Goldfarb's assertion to the effect that a society with the same number of poor but more rich people is more likely to have a Pareto optimal redistribution *available*, other things equal, hinges on an implicit assumption that *all potential recipients* must share in any transfers actually made. But whatever a community's size, a "rich" man, by earmarking his contributions, can bring about the same redistribution to any particular "poor" man that a community can accomplish vis-à-vis all poor men. This, of course, is what happens in families in which income recipients make continuing transfers to the other members.

IV. *Conclusion*

Our paper, as we have pointed out, questioned the neoclassical approach to distributional issues,[8] but it " . . . does not pretend to claim that fiscal reality does not deviate from the requirements of Pareto optimality" (Hochman and Rodgers, p. 556). We can, however, go beyond our original findings and suggest a number of other conclusions to which we have come in the course of our further thought on the subject. First, the logic of utility interdependence, as we have noted, implies that in-kind redistribution is in the Pareto optimal category. Second, so long as there are no legal or economic impediments to mobility among political jurisdictions, Pareto optimal transfers are the only kind of redistributive actions possible at the state and local jurisdictional levels. Third, the fact that individuals do accede to the social con-

[6] All this suggests that the notion of merit wants is redundant. The phenomena to which this generally refers are simply cases of particular-commodity utility interdependence. In such situations "gains from trade" are possible, and their realization may require collective action, just as national defense requires it. Musgrave himself recognizes that merit wants might best be explained in terms of interdependent utility in his Biarritz paper (see p. 143). Also see Buchanan (1960). The ideas discussed in this paragraph are discussed in more detail in Rodgers (1969).

[7] One of the authors deals with the N-person case in a similar way but in more detail than Goldfarb's comment. See Rodgers (1970).

[8] Another recent paper, which deals with the same question, is by S. C. Kolm.

tract and participate in the redistribution, despite their preference to free ride, suggests that redistributive transfers are a social good.

The question of how to decompose redistribution into its primary and secondary components and the development of theories capable of explaining each of these classes of transfers is worth further study, both at the theoretical and the empirical levels. For the theory our paper discussed to have practical relevance, research in this area must, among other things, determine whether utility functions are in fact interdependent and whether, where observed, such utility interdependence is benevolent, as our discussion has assumed. It appears, at this point, that the most fruitful way of beginning to do this is to examine the available data on private contributions to charity.[9]

[9] We have made a start in this direction. See our forthcoming article "Utility Interdependence and Income Transfers Through Charity."

REFERENCES

J. M. Buchanan, "The Theory of Public Finance," *South. Econ. J.*, Jan. 1960, *26*, 234–38.

——— and G. Tullock, *The Calculus of Consent*, Ann Arbor 1962.

H. M. Hochman and J. D. Rodgers, "Pareto-Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–57.

——— and ———, "Utility Interdependence and Income Transfers through Charity," in K. E. Boulding and M. Pfaff, eds., *Transfers in an Urbanized Economy: Theories and Effects of the Grants Economy* (forthcoming).

S. C. Kolm, "On the Optimal Production of Social Justice," in J. Margolis and H. Guitton, eds., *Public Economics*, New York 1969, 145–200.

M. S. March, "Public Programs for the Poor: Coverage, Gaps, and Future Directions," in S. J. Bowers, ed., *Proceedings of the National Tax Association*, New York 1967, 606–19.

S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quart. J. Econ.*, Feb. 1963, *77*, 95–112.

R. A. Musgrave, "Provision for Social Goods," in J. Margolis and H. Guitton, eds., *Conference On the Analysis of the Public Sector, Biarritz, 1966*, New York 1969, 124–44.

J. D. Rodgers, "On the Optimal Form of Income Transfers," Urban Institute Working Paper #1200–04, Washington 1969.

———, "Utility Interdependence and Income Redistribution," unpublished doctoral dissertation, Univ. Virginia 1970.

Substitution and Two Concepts of Effective Rate of Protection: Comment

By Herbert G. Grubel and Peter J. Lloyd*

In their recent paper in this *Review*, James Anderson and Seiji Naya have defined the effective rate of protection of industry $j(e_j)$ as:

$$e_j = \frac{v_j^d}{v_j} - 1 \tag{1}$$

where v_j^d and v_j are value-added per unit of output of industry j in equilibrium under protection and free trade, respectively. Free market value-added in turn is defined as

$$\begin{aligned} v_j &= p_j - p_i q_{ij} \\ &= p_j(1 - a_{ij}) \end{aligned} \tag{2}$$

where p_j and p_i are the free market prices of industry j's output and inputs, respectively, and q_{ij} and a_{ij} are free-market physical and value coefficients, respectively, of input i per unit of industry j's output.

Anderson and Naya then commit a serious error which invalidates the major proportion of their analysis. They argue that domestic value-added under protection can be *defined* in two different ways so that consequently there are two different definitions of e_j. In fact, however, there is only one definition of domestic value-added under protection:

$$\begin{aligned} v_j^d &= p_j(1 + t_j) - p_i(1 + t_i)q'_{ij} \\ &= p_j(1 + t_j)(1 - a'_{ij}) \end{aligned} \tag{3}$$

where t_j and t_i are the nominal tariff rates on output and input, respectively, and q'_{ij} and a'_{ij} are the coefficients already defined under protection.

Anderson and Naya argue that there are two definitions of e_j depending on whether it is assumed that $q'_{ij} = q_{ij}$ or whether $q'_{ij} \neq q_{ij}$ and they derive many complicated relationships for the two definitions of e_j. But it is clear that the *concept* of effective protection as defined in equation (1) is unchanged as a result of different assumptions about factor input substitution in the real world.

More precisely, the authors' mistake arises from the following procedure. They *assume* that $q_{ij} = q'_{ij}$, i.e., that factor substitution is zero, and obtain their equation 5 (not shown) which is a standard definition of effective protection.[1] They also assume a *CES* production function and derive equation 7 (not shown) which gives the precise functional relationship between a'_{ij} and a_{ij} and hence q'_{ij} and q_{ij}, in terms of the nominal tariffs and the elasticity of substitution. The invalid step in the analysis occurs when their equation 7, valid under the assumption that substitution does take place, is substituted into their equation 5, valid *only* under the assumption that substitution is zero, to derive their equation 8, which they claim is a measure of effective protections when substition is possible. The interpretation of formula 8 and the entire discussion of their second concept of effective protection under substitution therefore is meaningless.

What remains of their analysis can simply be summarized as follows. Assuming, as they do, that an industry's production function is *CES*, that factor prices are determined competitively, that we know the elasticity of substitution between the primary factor and the protected input (s_j), and we know the protected value input coefficient (a'_{ij}); then

$$a_{ij} = a'_{ij}\left(\frac{1 + t_j}{1 + t_i}\right)^{1-s_j} \tag{4}$$

This is their equation 7. Substituting equations (2)–(4) into (1) results in the following

* Associate professor of finance at the University of Pennsylvania and visiting fellow at the Australian National University. The second author is a Senior Research Fellow at the Australian National University.

[1] Equation numbers without parentheses refer to the original article.

measure of effective protection:

$$e_j = \frac{(1 + t_j) - (1 + t_i)a'_{ij}}{1 - a'_{ij}\left(\frac{1 + t_j}{1 + t_i}\right)^{1-s_j}} - 1 \tag{5}$$

which is the only valid expression providing an unbiased estimate of the effective rate of protection as defined in equation (1). Equation (5) could be expressed alternatively in terms of the free trade coefficient, a_{ij}, and is equivalent to Anderson and Naya's equation 10. From the equation, one can readily see that if it is assumed that $s_j=0$ when in fact $s_j>0$, then the estimate is unbiased if $t_i=t_j$. If, however,

$$a_{ij}[(1 + t_j)/(1 + t_i)]^{1-s_j} < 1$$

then the estimate is biased upward if $t_j>t_i$ and biased downward if $t_j<t_i$, as Anderson and Naya have shown by partial differentiation. Elsewhere we have presented the results of a computer simulation of equation (5) to indicate the sensitivity of the bias to changes in all parameters.

In the ambiguously worded last paragraph of Section II, Anderson and Naya appear to argue that the biases from making the false assumption about the factor substitution elasticity are different depending on whether one assumes one has knowledge of the protected (a'_{ij}) or the unprotected (a_{ij}) input coefficients. Simple transformation of equation (4) above to isolate a'_{ij} on the left-hand side and substitution of equations (2), (3) into (1) results in

$$e_j = \frac{(1+t_j)-(1+t_i)a_{ij}\left(\frac{1+t_j}{1+t_i}\right)^{s_j}}{1-a_{ij}} - 1 \tag{5'}$$

Inspection reveals that the ***direction*** of the biases from the wrong assumption that $s_j=0$ are the same whether one uses equation (5) or (5′). However, given identical numerical values for a_{ij} and a'_{ij} the ***extent*** of the bias from falsely assuming $s_j=0$ is not the same. This does not mean that the extent of the bias differs according to whether the free trade or protected version of the measure of effective protection is used. For any given s_j, the free trade and protected trade version of e_j are equivalent. These two statements are reconciled by noting that for any given free trade coefficient a_{ij} varying the value of s_j necessarily changes the equilibrium value of the protected trade coefficient and vice versa.

REFERENCES

H. G. Grubel and P. Lloyd, "Factor Substitution and Effective Tariff Rates," *Rev. Econ. Stud.*, forthcoming.

J. Anderson and S. Naya, "Substitution and two Concepts of Effective Rate of Protection," *Amer. Econ. Rev.*, Sept. 1969, *59*, 607–11.

Substitution and Two Concepts of Effective Rate of Protection: Reply

By JAMES ANDERSON AND SEIJI NAYA*

The possibility of substitution between primary and intermediate factors in the effective protection model raises a number of complex questions. Our paper skirted some of them, since it discussed measurement bias in two definitions of effective protection with substitution that were extant in the literature, but spent little time in analyzing their economic implications. A third definition of effective protection under *CES* production functions introduced by J. C. Leith was passed over entirely. In their comment, Herbert Grubel and Peter Lloyd question the logical possibility of one of our definitions and while their doubt is mistaken, it raises the broader issue of relating definitions of effective protection to the purpose of measuring effective rates in the first place. Therefore, we will first discuss the goal of effective protection measurement (whether the elasticity of factor substitution is zero or not), and then approach to the goal as an index of the desirability of various definitions of effective protection possible when substitution can occur. It will be seen that the criticized definition, discussed first by W. M. Corden, is possible and meaningful, and that it is far more desirable on theoretical and empirical grounds than the other definition we have considered. We will see that Leith's definition, on the other hand, may be preferable to the other two on theoretical grounds. Finally, we will clear up an evident confusion in "the ambiguously worded last paragraph" of Grubel and Lloyd's comment.

It is now evident, if there was ever any doubt, that effective protection is a partial equilibrium concept.[1] We wish, however, to use it to infer general equilibrium implications. Protection to an industry in general equilibrium can most usefully be defined in terms of output change, and so we would demand of an ideal measure of effective protection that a ranking of effective rates correspond to a ranking of general equilibrium percent output changes. This is an extremely harsh requirement, and the best that may be hoped for is an approach to the ideal. Output tariffs as a measure of protection may be thought of as the first step in this approach. Effective rates of protection were a second step that took account of tariffs as taxes on intermediate inputs. Under the original fixed coefficients assumption, effective rates were defined as the percent increase in partial equilibrium value-added per unit as a result of the tariff structure. The justification for this procedure is that in short-run (capital fixed) partial equilibrium, the measure will give rent due to the tariff structure as a proportion of free trade value-added. This rent, if positive, may then be used to bid for primary factors and thus expand (intermediate factors are assumed to be in infinitely elastic supply), and equal rent will induce greater expansion in a low value-added industry than in a high. Even if the fixed coefficients assumption is met, a number of difficulties prevent this measure from being an ideal one.[2] But violation of the fixed coefficient assumption as a source of distortion has received much attention. It is evident that of two industries with identical free trade coefficients and tariffs, the one with the greater ability to substitute against high priced inputs (whether primary or intermediate) will be able to expand more. Therefore the ideal measure ought to reflect this. If the definition adopted is our second one, partial equilibrium value-added per unit under pro-

* The authors are, respectively, assistant professor, Boston College and associate professor, University of Hawaii.

[1] See Anderson and Roy Ruffin for explicit general equilibrium treatments.

[2] See W. M. Corden, Harry Johnson, and Anderson for discussions.

tection as a percent of free trade value-added per unit, the true measure will be a decreasing function of the elasticity of substitution, given the free trade coefficients and a tariff constellation with the output tariff greater than average input tariffs—exactly the reverse of what is desired. This reflects the fact that protected equilibrium value-added per unit incorporates both the new higher prices of primary factors in the industry and the substitution that has occurred against them. Only if wage bills per unit were of any economic interest would the value-added definition be useful.

Therefore Corden rejected it and proposed instead an "impact" definition, which uses the non-equilibrium protected value-added per unit obtainable from combining protected prices with free trade coefficients. It gives the initial rent implicit in the tariff structure at free trade output and techniques as a proportion of free trade value-added. Contrary to Grubel and Lloyd, it is a perfectly meaningful and much more desirable definition of effective protection than the equilibrium value-added definition. It does have the undesirable property that greater ease of substituting against high priced inputs is not accounted for, but it is at least neutral in this regard. A third definition used by Leith is more promising in that it is an increasing function of the elasticity of substitution given the free trade coefficients and the tariff structure. Its weakness is that its theoretical economic meaning depends crucially on the specification of the production function. Leith defines effective protection as the percent change in the "price" of value-added due to protection. It is the same as the other two if no substitution is possible, but diverges from them if substitution can occur. In this latter case it may be thought of as rent to the primary factors as a group after substitution. The problem with it, as V. K. Ramaswami and T. N. Srinivasan have shown, is that unless substitution is neutral between intermediate and primary factors (all pairs of intermediate and primary factors have the same elasticity of substitution) the concept breaks down. In their counterexample with non-neutral substitution, they show it is no longer possible to identify rent to the primary group of factors; every primary factor is substituted against differently. In this case it may be better on theoretical grounds to utilize the impact concept, which still has a clear theoretical meaning, however crude. Of course, with non-neutral substitution, bias in its measurement is more complex than in our analysis.

Empirical research by Naya suggests that it doesn't make much difference in practice which definition is used, the rankings are scarcely affected. For example, a correlation between the usual estimate of effective protection and Leith's concept assuming an elasticity of substitution of 0.5 for 158 manufacturing sectors in Korea is .99. Also, a correlation between two assumed zero elasticity estimates using different years' input coefficients with the same tariff rates is .92 despite a considerable difference in these input coefficients. The average levels were different of course, but it is by now obvious that levels of protection are not particularly meaningful in themselves.

It remains to dispel a point in Grubel and Lloyd's last paragraph concerning bias in estimating the value-added concept. It is clearly stated in our paper that the false assumption of zero elasticity of substitution will give different estimates of the true value of the value-added measure, depending on whether the free trade or the protected coefficients are observed. The estimate using protected coefficients will always be larger. The true value of the effective rate is obviously the same whether we observe free trade or protected coefficients and thus the bias from making the assumption that $s_j=0$, as we stated, is greater in the protected situation. Grubel and Lloyd apparently concur in this. However, they also seem to be pointing out that if free trade (protected) coefficients are observed, there is a whole family of b_j (r_j) curves, one corresponding to each assumed elasticity of substitution. Only one corresponds to the correct elasticity, and only for this b_j (r_j) do the curves intersect at the true value of the effective rate under this definition. The others will intersect at the assumed s_j (as well as at $s_j=1$) and this

is what Grubel and Lloyd mean by saying that "for any given s_j, the free trade and protected trade version are equivalent." This tells nothing about the bias resulting from the substitution assumption, however, since a family of curves will be plotted as a function of s_j using wrongly estimated protected (free trade) coefficients for all assumed elasticities except the correct one. Only when the r_j and b_j with the correct coefficients are plotted can the bias from false substitution elasticity assumptions be discussed, and this was our procedure.

REFERENCES

J. Anderson, "General Equilibrium and the Effective Rate of Protection," *J. Polit. Econ.*, forthcoming.

——— and S. Naya, "Substitution and Two Concepts of Effective Rate of Protection," *Amer. Econ. Rev.*, Sept. 1969, *59*, 607–12.

W. M. Corden, "The Structure of a Tariff System and the Effective Protective Rate," *J. Polit. Econ.*, Dec. 1965, *73*, 573–94.

H. G. Johnson, "The Theory of Effective Protection and Preferences," *Economica*, May 1969, *36*, 119–38.

J. C. Leith, "Substitution and Supply Elasticity in Calculating the Effective Protective Rate," *Quart. J. Econ.*, Aug. 1968, *82*, 588–601.

S. Naya, 'The Pattern of Effective Protection Under Fixed Coefficients vs. Substitution: Based on the Korean Case," *mimeo.*

V. K. Ramaswami and T. N. Srinivasan, "Tariff Structure and Resource Allocation in the Presence of Factor Substitution," Discussion paper No. 33, July 1968, Indian Statist. Inst., Planning Unit, New Delhi, *mimeo.*

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Production," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

Investment Functions: Which Production Function?

By J. C. R. ROWLEY*

Recent studies of the determination of real fixed capital formation have been dominated by the approach suggested by Dale W. Jorgenson (1963, 1967). One deficiency of this approach is the need to specify an explicit form for the production function. This specification is not an integral feature of the neoclassical theory of optimal capital accumulation[1] but it is an essential prerequisite for econometric analyses. Jorgenson and his associated authors have always chosen the Cobb-Douglas form and they have argued that this specification is consistent with the results established in other areas of econometric research. Other economists have disputed the specification[2] and, in particular, they have suggested that the use of the Cobb-Douglas form must lead to an unjustified emphasis on the role of relative prices in the determination of investment expenditures. In this paper, an approximation is suggested whereby the alternative *CES* form can be fitted as a number of additional corrections to the final expression associated with the Cobb-Douglas form. Some numerical results are tabulated and these indicate that, for the British economy at least, the specification of a Cobb-Douglas form for the production function is inappropriate.

The symbols (K^*) and ($^*K$) are used to indicate those two expressions for "desired capital services" resulting when Jorgenson's approach is applied with two alternative specifications for the production function. The former corresponds to the Cobb-Douglas form whereas the latter is based on the *CES* specification. Additional notation is tabulated below.

X = Real output
K = Capital services
L = Labor input
p = Price of final output
q = Price of investment goods
c = "User-Cost" of capital input
ϕ = Adjustment factor representing the influences of tax provisions and a discount rate
I = Gross investment in real, non-dwelling, fixed assets
R = Replacement investment
u = Aggregate corporate tax-rate
v_1 = Aggregate rate of annual, or wear and tear, allowances
v_3 = Aggregate rate of initial allowances
v_4 = Aggregate rate of investment allowances
r = Dividend yield on industrial ordinary shares
δ = Index of capital longevity

The *CES* form of the production function can be written as the following expression.

$$(1) \qquad X^{\rho/\gamma} = \alpha_0 K^\rho + \beta_0 L^\rho,$$

where α, γ, α_0, and β_0 are given parameters. Differentiating partially, we obtain:

$$(2) \qquad \frac{\partial X}{\partial K} = \alpha_0 \gamma X^{(1-\rho/\gamma)} K^{(\rho-1)},$$

$$(3) \qquad K = (\alpha_0\gamma)^{1/(1-\rho)} \left(\frac{\partial X}{\partial K}\right)^{-1/(1-\rho)} \cdot X^{(1-\rho/\gamma)/(1-\rho)}$$

The following alternative expression would have resulted if we had used the Cobb-Douglas form.

* Assistant professor of economics at Queen's University. This topic was suggested by J. D. Sargan as part of an investigation into the determinants of real fixed capital formation within the postwar British economy.

[1] Jorgenson (1967) provides an outline of his neoclassical theory of investment behavior but his account contains no reference to the specification of a production function. His theoretical results are equally applicable for all functions satisfying certain convexity assumptions.

[2] See Robert Eisner and M. Ishaq Nadiri, Eisner, and Robert Coen.

(1a) $$X = MK^{\alpha}L^{\beta},$$

where M, α, and β are given parameters.

(3a) $$K = \alpha\left(\frac{\partial X}{\partial K}\right)^{-1} X$$

The neoclassical theory of optimal capital accumulation involves the maximization of net worth subject to production and replacement constraints. Jorgenson took a subset of the extremal conditions[3] indicated by this procedure and characterized them by a single optimality condition. This requirement implied the equality of the physical marginal productivity of capital services with a relative price-ratio formed from an implicit "user cost" of capital services and the price of final output. The expression for user cost is a product of the cost of investment goods and an adjustment factor representing characteristics of tax practice, physical deterioration in the capital input, and a discounting factor applied to future elements of net worth. Since tax practices differ between countries, the adjustment factor[4] suitable for American studies of real capital accumulation is different from that used in this study.

(4) $$\frac{\partial X}{\partial K} = \frac{c}{p} \quad \text{(Optimality Condition)}$$

(5) $$c \equiv q\left(\frac{1}{\phi}\right)$$

(6) $$\phi \equiv \frac{1 - u}{(1 - uv_3 - uv_4)(\delta + r) - uv_1\delta}$$

Substitution of the imputed price-ratio from the optimality condition (4) for the marginal product in (3) provides the expression for desired capital services associated with the *CES* form of the production constraint.

[3] My discussion paper outlines Jorgenson's approach and indicates suitable modifications for the British economy.

[4] The derivation of equation (6) is explained in Rowley. This expression should be contrasted with those indicated by Coen, and by Hall and Jorgenson.

(7) $$(^*K) = \alpha_1\left(\frac{p\phi}{q}\right)^{m_1} X^{m_2}$$

where

$$m_1 = \frac{1}{1 - \rho}$$

$$m_2 = m_1\left(1 - \frac{\rho}{\gamma}\right)$$

and

$$\alpha_1 = (\alpha_0\gamma)^{m_1}$$

In the special case specifiying constant returns to scale for the production function, γ is unity and, hence, m_2 is unity.

In the general case,

(8) $$\rho = \frac{m_1 - 1}{m_1} = \frac{n_1}{n_1 + 1}$$

and

(9) $$\frac{1}{\gamma} = \frac{m_1 - m_2}{m_1 - 1} = \frac{n_1 - n_2}{n_1}$$

if we define

$$n_i \equiv m_i - 1 \qquad \text{for } i = 1, 2$$

An expression for the alternative (K^*) can be obtained by a similar substitution from (4) to (3a).

(10) $$(K^*) = \alpha\left(\frac{p\phi}{q}\right)X$$

The elasticity of substitution, σ, is equal to m_1, or (n_1+1). As m_1 tends to unity (as n_1 tends to zero), the *CES* form may be replaced by the Cobb-Douglas form and $(^*K)$ replaced by (K^*). We can consider the equation (7) as a definition of a family of concepts of desired capital services generated by setting the three parameters at different levels. The choice of the production function imposes constraints on "suitable" members of this family. Jorgenson's specification of the Cobb-Douglas form leads directly to the implication that the elasticities of desired capital services with respect to relative prices and output have a common

TABLE 1

Specification	n_1	n_2
Cobb-Douglas	0	0
CES	$(\sigma-1)$	NONE
CES with constant returns to scale	$(\sigma-1)$	0

value, unity. This is the feature of Jorgenson's work that has attracted most criticism. Define the following measure of desired capital services:

$$(11)\qquad \begin{aligned}(^*K^*) &\equiv \alpha_1\left(\frac{p\phi}{q}\right)X\\ (^*K) &= (^*K^*)\left(\frac{p\phi}{q}\right)^{n_1}X^{n_2}\end{aligned}, \quad \text{so that}$$

If n_1 and n_2 are sufficiently close to zero, we can use the following approximation.[5]

$$(12)\qquad \begin{aligned}&\left(\frac{p\phi}{q}\right)^{n_1}X^{n_2}\\ &\qquad \simeq 1 + n_1\,log_e\left(\frac{p\phi}{q}\right) + n_2\,log_e X\\ &(^*K) \simeq (^*K^*)\left[1 + n_1\,log_e\left(\frac{p\phi}{q}\right)\right.\\ &\qquad\qquad \left. + n_2\,log_e X\right]\end{aligned}$$

Let ∇ be the difference operator defined, for an arbitrary variable z_t, by the equality of ∇z_t and $(z_t - z_{t-1})$.

$$(13)\qquad \begin{aligned}\nabla(^*K) &\simeq \nabla(^*K^*)\\ &\quad + n_1\nabla\left[(^*K^*)\,log_e\left(\frac{p\phi}{q}\right)\right]\\ &\quad + n_2\nabla[(^*K^*)\,log_e X]\end{aligned}$$

Table 1 contains a summary of the prior restrictions imposed on the coefficient of this approximation (13) by the choice of production function and by the specification of constant returns to scale. A desirable feature of the approximation is its convenience for testing the validity of these prior con- constraints using conventional statistical methods.

Some Empirical Results

A second characteristic of Jorgenson's work is the use of lags generated by rational functions of the lag-operator. Net investment is to be explained by the product of a weighting function and the temporal stream of changes in desired capital services. The empirical results tabulated below were obtained for one specification of the weighting function. This specification was chosen after an analysis[6] of the model using the Cobb-Douglas form for the production constraint. It was indicated by conventional statistical measures and by the standard-error of estimate criterion advocated by Jorgenson and Stephenson.[7] We shall use θ to represent the lag-operator defined, for an arbitrary variable Z_t, by the equality of θZ_t and Z_{t-1}. $W(\theta)$ represents the weighting function.

$$(14)\qquad W(\theta) = \frac{w_1\theta^3 + w_2\theta^4}{1 - b\theta}, \qquad W(1) = 1$$

$$(15)\quad (I-R)_t = W(\theta)\cdot\nabla(^*K)$$

$$(16)\quad \begin{aligned}(I-R)_t &= b(I-R)_{t-1}\\ &\quad + \sum_{j=3}^{4} a_{1,j-2}\nabla(^*K^*)_{t-j}\\ &\quad + \sum_{j=3}^{4} c_{1,j-2}\nabla\left(^*K^*\,log_e\frac{p\phi}{q}\right)_{t-j}\\ &\quad + \sum_{j=3}^{4} c_{2,j-2}\nabla(^*K^*\,log_e X)_{t-j}\end{aligned}$$

where

$$\begin{aligned} a_{11} &= w_1; & a_{12} &= w_2\\ c_{11} &= w_1n_1; & c_{12} &= w_2n_1\\ c_{21} &= w_1n_2; & c_{22} &= w_2n_2\end{aligned}$$

One difficulty with (16) is the presence of the unknown scalar α_1 in $(^*K^*)$. Since this

[5] An account of this approximation is provided in the Appendix.

[6] See Rowley for a tabulation of results for many different specifications of the weighting function.

[7] Also Eisner and Nadiri.

TABLE 2—ADDITIONAL REGRESSORS

Variable	(a)	(b)	(c)	(d)	(e)	(f)
$(I-R)_{t-1}$	.6156	.3868	.3991			
	5.9564	3.0589	3.0088			
$\nabla({}^*K^*)$						
t-3	.2938	.4346	.4258	.6459	.6643	.6646
	4.1412	5.2062	4.8441	11.3161	16.0389	15.4675
t-4	.2535	.3987	.3925	.6459	.6386	.6400
	3.4211	4.6191	4.3473	11.0163	15.6030	15.0706
$\nabla({}^*K^* \log p\phi/q)$						
t-3		−.0614	−.0816		−.1013	−.0952
		−2.0157	−1.4744		−3.2425	−1.5138
t-4		− .0666	− .0481		− .1204	− .0990
		−2.0102	− .8119		−3.7820	−1.5313
$\nabla({}^*K^* \log X)$						
t-3			.0018			− .0006
			.4296			− .1267
t-4			− .0019			− .0017
			− .4437			− .3728
$100\bar{R}^2$	91.82	93.57	93.67	82.16	91.42	91.47
$1+(c_{11}/a_{11})$		.859	.808		.848	.857
$1+(c_{12}/a_{12})$		.846	.877		.811	.845
$1+(c_{21}/a_{11})$			.996			.999
$1+(c_{22}/a_{12})$			.995			.997

scalar is multiplicative, the least squares fit for (16) will yield estimates of b and the products of α_1 and each of the remaining parameters. The equation was fitted to quarterly data for the British economy extending, for the dependent variable, from 1958 to 1965 and Table 2 contains these parametric estimates for various prior restrictions. Each cell in the major section of the table contains a parametric estimate and its corresponding Student's t-statistic. The over-identifying restrictions were ignored in the first stage of the estimating procedure but were used to derive estimates of m_1 and m_2 subsequently. These secondary estimates are contained in the lowest section of the table.

Results for six equations are presented. (a) and (d) embody the Cobb-Douglas form for the production function whereas the remainder embody the alternative *CES* form. (b) and (e) are based on the assumption of constant returns to scale. Three substantive conclusions may be drawn from the table.

(1) The most suitable specification for the production function in this context is the *CES* form with constant returns to scale.

(2) The use of the Cobb-Douglas function leads directly to an exaggeration of the importance of relative prices in the determination of investment. Appropriate elasticities of desired capital services to output and relative prices are approximately unity and 0.85, respectively.

(3) The arguments of the relative-price variable were collectively very significant during the period under review, even though less significant than would be indicated by the Cobb-Douglas model.

APPENDIX

$$\log_e (1-s) = -s - \frac{s^2}{2} - \frac{s^3}{2} - \dots \quad \text{for } -1 \leq s < 1$$

Define $r = 1 - s$

$$\log_e r = -(1-r) - \frac{(1-r)^2}{2} - \frac{(1-r)^3}{3}$$

$$\text{for } 0 < r \leqq 2$$

For r sufficiently close to unity (s sufficiently close to zero), we can use the following approximation.

$$\log_e r \simeq -(1-r)$$

$$r \simeq 1 + \log_e r$$

Suppose $r = \Pi^n_{j=1} x_j^{\alpha j}$ for some variables $\{x_j\}$
Then,

$$\left(\prod_{j=1}^{n} x_j^{\alpha j}\right) \simeq 1 + \log_e \left(\prod_{j=1}^{n} x_j^{\alpha j}\right)$$

$$= 1 + \sum_{j=1}^{n} \alpha_j \log_e x_j$$

provided each $\{x_j\}$ is sufficiently close to zero. Here sufficiency depends on the closeness of each $\{x_j^{\alpha}{}_j\}$ to unity.

$$\prod_{j=1}^{n} x_j^{1+\alpha j} = \left(\prod_{j=1}^{n} x_j\right)\left(\prod_{j=1}^{n} x_j^{\alpha j}\right)$$

$$\simeq \left(\prod_{j=1}^{n} x_j\right)\left(1 + \sum_{j=1}^{n} \alpha_j \log_e x_j\right)$$

$$= \prod_{j=1}^{n} x_j + \left(\prod_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} \alpha_j \log_e x_j\right)$$

$$= \prod_{j=1}^{n} x_j + \sum_{j=1}^{n} \alpha_j \left(\prod_{i=1}^{n} x_i\right) \log_e x_j$$

REFERENCES

R. Coen, "Tax Policy and Investment Behavior: Comment," *Amer. Econ Rev.*, June 1969, *59*, 370–79.

R. Eisner, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 379–88.

——— and M. I. Nadiri, "Investment Behavior and the Neo-Classical Theory," *Rev. Econ. Statist.*, Aug. 1968, *50*, 369–82.

R. E. Hall and D. W. Jorgenson, "Tax Policy and Investment Behavior," *Amer. Econ. Rev.*, June 1967, *57*, 391–414.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 247–59.

———, "Anticipation and Investment Behavior," *The Brookings Quarterly Econometric Model of the United States*, ch. 2, Chicago 1965.

———, "The Theory of Investment Behavior," Contribution to *Nat. Bur. Econ. Res. Conference "Determinants of Investment Behavior*, 1967, *35*, (2), 169–220.

——— and James Stephenson, "Investment Behavior in U.S. Manufacturing, 1947–1960," *Econometrica*, Apr. 1967, *35*, 169–220.

J. C. R. Rowley, "Fixed Capital Formation in the British Economy, 1956–1965," Discussion Paper No. 10, November 1969, Inst. of Econ. Res., Queen's Univ., Kingston.

Transport Costs, Tariffs, and the Pattern of Industrial Protection

By W. G. WATERS II*

This study examines the protection provided by international transport costs for *U.S.* industries using the 1958 input-output table. The level and pattern of protection among industries is examined both for tariffs and freight factors (freight and insurance charges relative to f.o.b. value of imports). Calculations are made both on a nominal and effective rate basis.

I. *Transport Costs and the Theory of Effective Protection*

For the sake of brevity, it is assumed that the reader is acquainted with the theory of effective protection, hence the underlying assumptions and details will not be discussed.[1] The effective tariff is the percentage change in value-added[2] in an industry made possible by a tariff structure. The effective tariff does *not* measure protection to domestic primary resources vis-à-vis those of the rest of the world. All that is considered is the difference in value-added contributed by domestic primary resources with and without a tariff structure.

Extension of the theory of effective protection to include transport costs is straightforward.[3] The existence of freight costs causes divergence between domestic and foreign prices hence a difference in domestic value-added from what it would be in a world without transport costs.

The effective tariff is more important information since tariffs are an instrument of commercial policy whereas, by and large, transport costs are not. A comparison of effective tariffs is a guide to the distortion in resource allocation between the case of protection and that of free trade. An effective freight factor can be viewed as a guide to the "natural effect" on resource allocation which results from the existence of a structure of freight factors, i.e., the change in resource allocation between a free trade and a frictionless world case.

One assumption concerning the nominal freight factors bears mention. It is useful to assume that the freight factor for any product is the same regardless of the geographic dispersion of trading partners. This is not too unreasonable. Freight costs do vary among routes and commodities but such variation tends to be quite small relative to the variation in value per weight unit of different commodities, e.g., high valued products have low freight factors not because their costs of transport per weight unit are less than for low valued products,

* The University of British Columbia. I would like to thank, without implicating, Robert E. Baldwin, James E. Anderson, Larry J. Wipf and members of the stimulating Workshop in International Economics at the University of Wisconsin. Financial assistance was primarily provided by the Ford Foundation via a research fellowship administered by the department of economics at the University of Wisconsin. Necessary computer time was furnished by the Graduate Research Committee at the University of Wisconsin. Revisions were carried out while on a postdoctoral fellowship sponsored by the Canadian Transport Commission. While the "inputs" provided by the aforementioned are hereby gratefully acknowledged, the responsibility for the resultant "output" is mine alone.

[1] The literature is extensive, but one might select Bela Balassa (1965), W. M. Corden, and Harry Johnson (1965) as basic references.

[2] Corden has correctly pointed out that it is impossible to distinguish between the contribution of domestic primary resources (value-added) and that of domestic nontraded intermediate inputs. Except for one instance, value-added is used as the base for computing our effective rates although it is conceptually incorrect. Its use permits direct comparisions to be made between our calculations and others that have been made for the United States, specifically Balassa (1965) and Giorgio Basevi.

[3] Balassa (1968) and Johnson (1969) include a formula for computing a "natural rate of protection" although they do not pursue it empirically. See also my dissertation.

A complete derivation of the various effective rates is in a mathematical appendix available from the author upon request.

in fact they tend to be higher, but because the per unit value is so much higher.[4]

One further issue is that of substitution among inputs and factors. Our rates are calculated using the Corden formulation ((1) below) which is an over-estimate if the elasticity of substitution is not zero.[5] But this applies to effective rates for both tariffs and freight factors and since a major interest is the pattern of protection with and without consideration of transport costs, a unidirectional bias is not expected to cause great anguish. In any case, an additional calculation can be made under the alternate assumption of a unitary elasticity of substitution.

An effective tariff for j, (et_j) can be calculated by

$$et_j = \frac{t_j - \sum_{i=1}^{n} a_{ij} t_i}{1 - \sum_{i=1}^{n} a_{ij}} \tag{1}$$

$$et'_j = \frac{t_j - \sum_{i=1}^{n} a'_{ij} t_i}{1 - \sum_{i=1}^{n} a'_{ij}} \tag{2}$$

where t_j is the tariff on final product, t_i the tariff on intermediate input i, a_{ij} the free trade Leontief coefficient, and a'_{ij} the similar coefficient under protection. Equation (1) is derived assuming a zero elasticity of substitution among all inputs and factors. Equation (2) is a measure *before* resources move in response to protection, under the assumption of a unitary elasticity of substitution.[6] Effective rates for tariffs, freight factors and their combination will be calculated in both forms. If in equations (1) or (2), we substitute a freight rate factor for t, we would obtain an *effective* freight factor, similar in principle to an effective tariff.

II. *The Data and Empirical Results*

The tariff rates are estimated from basic data for the input-output table; they are defined as the ratio of actual duties paid relative to the value of competitive imports.[7] These "import-weighted" tariffs are an underestimate of the actual level of protection for the familiar reason that only imports which successfully enter the country in spite of tariffs are recorded. An important reason for our use of import-weights is that our freight factors, by necessity, are estimated in the same manner, i.e., actual freight and insurance margins relative to the value of transferred (competitive) imports.

One caveat applies: the use of an input-output table is not without disadvantages, e.g., the unavoidably high level of aggregation of each industrial classification. This potential problem of heterogeneity within classifications should caution against interpreting the results too precisely for the individual industries.

Nominal tariffs and freight factors were computed for fifty-seven of the industry classifications in the input-output table. Other classifications were excluded since they refer to services rather than importable commodities or because of data problems of one form or another.[8] A list of the various rates

[4] There is evidence in support of this, see Herman Karreman (p. 16) and Carmellah Moneta (p. 51).

Canada, the *U.S.* major trading partner, is probably an exception (my freight margins consist only of estimates of air and ocean freight costs). Freight factors may be notably less on Canadian imports. This suggests that the structure of freight factors is also an over estimate of the natural protection for Canadian suppliers to the *U.S.* market vis-à-vis non-North American suppliers.

[5] See Anderson and Naya (p. 608–9) and Corden (pp. 233–35). For actual estimation, (1) must be expressed in terms of the observable coefficients (a_{ij}') rather than the unobservable free trade coefficients.

[6] Corden and Anderson and Naya.

[7] In practice the United States imposes its tariffs on foreign port values (f.o.b.).

[8] Omitting the service classifications leads to a serious under estimate of the importance of transport costs. "Services" may be interpreted as products which are non-traded because they encounter very high freight factors.

Household Furniture and Other Furniture and Fixtures were excluded because the imports (and duties and freight) are directly allocated in the *I-O* table rather than treated as competitive imports. Coal Mining and Household Appliances are excluded because transport and duty figures, respectively, were not available.

TABLE 1—TARIFFS AND FREIGHT FACTORS

	Nominal Tariff	Nominal Freight	Total Nominal	Effective Tariff	Effective Freight	Total Effective	Effective Tariff-1	Effective Freight-1	Total Effective-1
Lowest Value	0	1.16	1.66	− 9.28	−12.75	−20.25	−11.42	−21.13	−32.72
Highest Value	41.37	41.73	50.81	112.69	105.89	146.16	76.76	74.12	92.70
Range	41.37	40.57	49.15	121.97	118.64	166.41	88.18	95.25	125.42
Median	11.25	6.31	19.20	16.99	6.54	26.95	16.85	6.24	25.95
Unweighted Mean	11.12	9.18	21.13	19.80	13.40	35.21	17.10	11.82	30.11
Standard Deviation	9.01	7.96	10.98	24.97	21.70	33.54	18.25	18.07	25.15
Coefficient of Variation	.81	.87	.52	1.26	1.62	.95	1.07	1.53	.84
Weighted Mean	6.23	10.02	16.79	9.56	13.08	23.88	8.50	13.89	23.07
Weighted Standard Deviation	7.02	7.11	9.65	20.46	14.18	27.21	14.90	16.79	23.10
Weighted Coefficient of Variation	1.13	.71	.57	2.14	1.08	1.14	1.75	1.21	1.00

of protection are in a table which will be supplied by the author upon request. The major point to be made is that freight factors are, in general, not trivially small relative to tariffs. They have a comparable level, range and dispersion (see Table 1). However, the correlation between tariffs and freight factors is low (Table 2), in fact negative although this is not significant at the 5 percent level of confidence. Total nominal protection has a greater level (median and weighted mean values of 19.20 and 16.79) and a slightly greater range. But neither the nominal tariffs nor nominal freight factors are really strong guides to the pattern of protection given by the total nominal rates (correlation coefficients are .674 and .542, respectively).

The range of the effective tariffs is approximately three times the range of the corresponding nominal tariffs while the median level is about one and one-half times the corresponding median nominal rate. These results are consistent with the findings for the United States of both Balassa (1965 p. 588) and Basevi (pp. 154, 156). A comparison between nominal and effective freight factors is similar to the comparison between nominal and effective tariffs except that the distribution of effective freight factors is skewed downward (the median value is noticeably less than the unweighted mean).

The coefficient of correlation between the effective tariffs and effective freight factors is −.121. This is not significantly different from zero, although it is significantly less than 1.0. As can be seen by the coefficients of correlation in Table 2, the pattern of protection revealed by consideration of total protection on an effective basis, although positively related, is quite different than the pattern given by either effective tariffs or effective freight factors.

Also of interest is the correlation between effective rates and the corresponding nominal ones. For each of these three instances, the correlations are extremely high. The Spearman rank correlation coefficient between nominal and effective tariffs is .9575, that between nominal and effective freight factors is .8992 while that for total nominal and total effective is .9454. Although calculation of an effective rate is intended to do more than just provide an ordering of industrial protection (i.e., provide a guide to the relative impact on resource allocation due to the tariff structure), it appears that the much more easily obtained nominal rates might do a satisfactory job of indicating the pattern of protection among industries. The correlation between the nominal and effective tariffs reported by Basevi and Balassa further confirms this finding. The rank correlation coefficients for the United States data (1958) are .8890 and .8407, respectively.[9] This should not come as too great of a surprise; if a tariff structure is escalated in a fairly uniform manner, then the nominal

[9] Correlation coefficients computed by this author.

TABLE 2—CORRELATION BETWEEN VARIOUS RATES OF PROTECTION

	Nominal Tariff	Nominal Freight	Total Nominal	Effective Tariff	Effective Freight	Total Effective	Effective Tariff-1	Effective Freight-1	Total Effective-1
Nominal Freight	−.1746 −.256								
Total Nominal	.5777 .674	.5386 .542							
Effective Tariff	.9575 .952	−.1276 −.199	.5924 .674				.9977 .988		
Effective Freight	−.1073 −.173	.8992 .933	.5721 .565	−.0412 −.121				.9929 .946	
Total Effective	.5302 .633	.5422 .491	.9454 .928	.6023 .713	.6609 .607				.9854 .955

Note: The upper figure appearing in each element is the Spearman's Rank Correlation Coefficient. The lower figure in each element is the coefficient of correlation.

rates would be expected to be a reliable guide to the pattern revealed by effective rates.[10]

Table 1 in its last three columns, also lists the level and dispersion of the effective rates computed under the assumption of a unitary, rather than zero, elasticity of substitution. The level and range of this set of rates are less than the previous calculations. While this new assumption concerning the elasticity of substitution brought some changes in the level and dispersion of effective rates, the pattern among industries was largely unchanged. The rank correlation coefficient between the two effective tariffs is .9977. The same coefficients between effective freight factors and total effective rates are .9929 and .9854, respectively.[11] This finding is valid only if the elasticity of substitution is the same for all industries. If it is not and varies significantly (as it could), then it is uncertain how much reliability can be assigned to any group of effective rates regardless of how they are computed. However, freight factors and tariffs are not strongly correlated with one another. Even if the elasticity of substitution in all industries were known, and "true" effective tariffs were computed, the pattern of protection thus revealed would probably still be different than a pattern revealed with a consideration of transport costs as well.

III. *Further Results and Implications*

I am reluctant to attach too much faith to the exact values computed for various industries, because of the gross level of aggregation in this study relative to other studies (specifically Basevi). However, following Balassa's example it is possible to group our commodity classifications into four larger groups comprised of consumer goods and investment goods along with the two types of intermediate products:

> Semi-manufacturers whose main inputs are natural raw materials have been classified as intermediate products I, while all intermediate goods at higher levels of fabrication have been included in intermediate products II. [Balassa 1965, p. 591]

The grouping scheme used by Balassa was followed as closely as possible although numerous cases did not fit easily into his grouping. Following Balassa's example the input-output columns which show the al-

[10] Although Corden notes that "the order (of effective rates) is likely to be quite different from a similar scale based on nominal tariff rates. . ." (p. 224).

[11] J. Clark Leith examined the impact on effective rates of altering substitution and supply elasticities. Although correlation coefficients were not reported, the ranking of effective rates appears to be hardly affected by changing the elasticity of substitution between inputs and factors (p. 598–99).

location of industry output to intermediate and final demands were used to assist in grouping the product classifications.

Table 3 lists these results along with Balassa's values for the United States. Different levels of aggregation and different weighting procedures suggest that our results could not be expected to be precisely the same. The only notable difference, however, is that our nominal tariff for Intermediate Products I is somewhat less (about one-half) than Balassa's and our effective tariff quite a bit less. In general, our nominal and effective tariffs appear to have about the same pattern among industry groups as Balassa's rates. Tariffs tend to be higher on the more highly processed industries, i.e., the tariff structure is escalated.

The total protection for Intermediate group I is of considerable interest. Despite the fact that it has a much lower effective tariff, when transport costs are included the total protection climbs to a level commensurate with the level of total protection for the other product groups. This is true for both total nominal and total effective rates.

TABLE 3—WEIGHTED AND UNWEIGHTED MEANS FOR ALL INDUSTRIES COMBINED AS WELL AS INDUSTRY GROUPS[c]

	Nominal Tariffs	Nominal Freight Factors	Total Nominal	Effective Tariffs	Effective Freight Factors	Total Effective
All Sectors[a]						
Unweighted mean	11.12	9.18	21.13	19.80	13.40	35.21
Median	11.25	6.31	19.20	16.99	6.54	26.95
Weighted mean	6.23	10.02	16.79	9.56	13.08	23.88
Balassa's weighted mean	11.6	—	—	20.0	—	—
Intermediate Type I[b]						
Unweighted mean	3.37	15.06	18.91	4.34	23.71	29.17
Median	3.43	12.19	17.12	3.75	20.02	30.42
Weighted mean	2.45	13.59	16.42	1.77	20.21	22.66
Balassa's weighted mean	8.8	—	—	17.6	—	—
Intermediate Type II[b]						
Unweighted mean	11.60	9.19	21.75	22.11	14.45	39.22
Median	11.16	6.26	19.01	17.33	6.16	23.98
Weighted mean	8.51	7.96	17.13	15.75	8.51	26.16
Balassa's weighted mean	15.2	—	—	28.6	—	—
Consumer Goods[b]						
Unweighted mean	18.41	6.22	25.67	33.86	8.01	43.81
Median	14.21	5.99	22.93	25.75	6.80	33.54
Weighted mean	10.66	6.77	18.11	15.31	8.05	24.36
Balassa's weighted mean	17.5	—	—	25.9	—	—
Investment Goods[b]						
Unweighted mean	10.94	5.91	17.54	16.17	5.00	22.30
Median	11.98	5.72	18.15	17.48	4.41	25.55
Weighted mean	7.46	5.01	12.99	7.53	3.03	11.78
Balassa's weighted mean	10.3	—	—	13.9	—	—

[a] The number of sectors in each category are 57, 10, 28, 9, and 10, respectively.

[b] The four categories of goods are identified in the table of freight factors which the author will make available upon request.

[c] Balassa weighted his means by the combined values of imports for all countries in his study. Our means are weighted by actual *U.S.* imports.

TABLE 4—MEAN RATES OF PROTECTION (CORDEN TYPE) WEIGHTED BY INDUSTRY SHARE IN TOTAL VALUE-ADDED PLUS NONTRADED INPUTS[a]

	All Industries	Intermediate Products I	Intermediate Products II	Consumer Goods	Investment Goods
Effective Tariff	10.41	2.88	12.56	10.20	12.60
Effective Freight	7.30	13.34	6.88	6.24	3.16
Total Effective Rates	18.47	16.61	20.30	17.09	16.15

Note: The tariffs used for these calculations are not the same as those used elsewhere in the paper. As a result, the level of these rates cannot be directly compared with the level of other rates.

[a] Value-added for Intermediate Products I is 15 per cent of total value-added. Corresponding figures for the remaining groups are 48, 27, and 10 percent, respectively.

These results suggest that once the protection of transport costs is considered, it is no longer readily apparent that protection is higher for more highly processed industries. The escalation in the tariff structure among these groups appears to be largely offsetting deescalation in the freight factors. The latter implies that trade in processed goods relative to less or nonprocessed ones would be encouraged under free trade. Under escalated tariff structures, the opposite tendency exists.

As frequently mentioned, the effective rates are interpreted as a guide to the direction resources will tend to flow due to the existence of a tariff structure (or a set of freight factors). A rough estimate of the relative impact of tariffs vis-à-vis freight factors can be obtained by comparing the mean level of these rates where the weights are the industries' share of total value-added plus nontradeable inputs.[12] These are listed in Table 4. The mean effective tariff for all commodities, 10.41, is slightly greater than the corresponding mean effective freight factor 7.30. The coefficients of variation are about the same (1.11 and 1.17, respectively) so the variability among industries does not appear to be significantly different. Our interpretation is that the impact on resource allocation resulting from *U.S.* tariff policy is probably greater than the impact of freight factors.[13] Bear in mind that the structure or pattern among industries is somewhat different. This can be seen by looking at the rates weighted by value-added for the four subgroups. The combined impact of tariffs and freight factors (total effective rates) yields a noticeably higher level of protection although they do not differ among industry groups as much as tariffs considered by themselves.

An alternative approach is to estimate the change in *U.S.* imports that would accompany the elimination of either form of protection. A simple partial equilibrium estimate of the percentage change in imports dM/M′ which would result if tariffs (or freight factors) were simultaneously removed for all countries is given by[14]

$$\frac{dM}{M'} = \sum_{k=1}^{4}\left[\left(\eta_k \frac{C_k'}{M_k'}\,\frac{\bar{t}_k}{1+\bar{t}_k}\right) - \left(e_k \frac{X_k'}{M_k'}\,\frac{\overline{et}_k}{1+\overline{et}_k}\right)\right]\frac{M_k'}{M'} \tag{3}$$

[12] For this test, a new set of effective rates were computed using the more correct base (see fn. 2, *supra*).

The weights are the industries' share of total free trade value-added plus non-tradeables for tariffs and share in total "frictionless trade" value-added plus non-tradeables for freight factors.

[13] However, the mean freight factor is seriously underestimated because we have not included nontraded products, i.e., those which, by definition, have such high freight factors to exclude them from trade.

[14] We must also assume the cross-elasticities of demand can be safely ignored.

Equation (3) omits consideration of the change in imports for intermediate use. If we assume similar production functions and recognize that there are both imports and exports for nearly every industrial classification in an input-output table, then we might expect offsetting increases in imports and exports for intermediate use when all countries remove tariffs simultaneously. Cf. Balassa (p. 590).

TABLE 5—MEAN RATES OF PROTECTION ON PROCESSED GOODS[a]

	Nominal Tariff	Nominal Freight	Total Nominal	Effective Tariff	Effective Freight	Total Effective
Processed goods of interest to *LDCs*	19.42	6.07	26.62	40.39	8.94	52.53
Other Processed Goods	10.97	9.36	21.26	19.25	14.35	35.82

[a] Processed goods are Intermediate Type II goods and consumer goods (see Table 4, *supra.*). Processed goods of interest to *LDCs* are identified in the appendix which the author will supply upon request.

The subscript k refers to the four product-groups; n_k and e_k are the elasticities of demand and domestic supply, respectively; C_k'/M_k' and X_k'/M_k' are protected consumption and domestic production relative to imports; and the weights M_k'/M' are the proportion of imports for each group relative to total imports. The tariffs $\bar{t}_k$ and $\overline{et}_k$ are the medians reported in Table 3.

Assuming some representative values for the unknown variables in equation (3) Balassa calculated that eliminating tariffs among all countries would lead to a 38.2 percent increase in *U.S.* imports.[15] A similar calculation based on our tariffs and import-weights indicated a 49.70 percent increase.[16] If freight costs were eliminated among countries (replace $\bar{t}_k$ and $\overline{et}_k$ in (3) by the corresponding nominal and effective freight factors) *U.S.* imports would rise by 21.75 percent. The latter is an underestimate relative to that for tariffs since no allowance was made for noncompetitive imports nor for possible trade in what are non-tradables at present.

Rates of protection are not solely of interest to the domestic economy, they are also of interest to foreign countries. We might briefly consider the level of protection (both of tariffs and freight factors) on products of interest to less developed countries (*LDCs*). Table 5 lists the unweighted mean levels of protection on a group of processed good industries (Intermediate Type II goods and consumer goods) which are of special interest to *LDCs*. These include textiles and textile products, hides and leather products, and miscellaneous manufacturing. Also calculated are the means for the remaining Intermediate Type II and consumer goods. The figures indicate that tariffs (both nominal and effective) do tend to be noticeably higher on the products of special interest to *LDCs*.[17] The freight factors are significantly lower than the tariffs and appear to be slightly lower than those for other processed goods. The total nominal and total effective rates are higher for the products of interest to *LDCs* than those for the other processed goods, but not in the same proportion as the difference between tariffs considered alone.[18]

The total nominal rate 26.62 implies that the delivered price from a competitive *LDC* supplier will be 26.62 percent above his costs of production (measured in dollars). If we would assume identical production functions and input prices in the United States and *LDCs*, this would imply that the returns to United States immobile, primary resources can be 52.53 percent higher than the returns to the corresponding *LDC* primary resources and inputs.

IV. *Summary*

This paper has explored the protection effect of international transport costs relative

[15] The elasticities of demand for the four product groups beginning with Intermediate I were −.2, −.3, −1.0, and −.3 while elasticities of supply were .1, .2, .8, and .3. The ratio of consumption and production relative to imports were assumed equal and assigned a value of 4 for all industries, the latter based on an earlier argument by Floyd. Also see Balassa (1965 p. 592–93).

[16] Balassa's own rates with our weights yielded a 58.4 percent increase.

[17] Balassa (1968) reports similar findings in his much more detailed analysis.

[18] Omitting Footwear and other leather products, which has a very high nominal and effective tariff, does not noticeably alter our findings. The mean nominal and effective tariffs reduced to 16.98 and 32.81, respectively.

to tariffs, both on a nominal and effective basis, for the United States. We find that excluding transport costs from consideration not only understates the level of protection for domestic vis-à-vis foreign resources, but also significantly alters the pattern of protection among industries. The escalation in the United States' tariff structure (the increased tariff levels on more processed goods) is partially offset by deescalation in the freight factors.

Some other empirical observations are: first, there appears to be a strong correlation between the nominal and effective rates for the United States (whether for tariffs, freight factors or their combination). That is, the pattern of protection among industries revealed by nominal rates is largely the same as that given by the more difficult to obtain effective rates. This property is also borne out in rates of protection computed by Balassa and Basevi for the United States.

A second finding of general relevance to the theory of effective protection is that altering the assumed elasticity of substitution from zero to one brought negligible change in the pattern of industrial protection given by rank correlations. This finding is a very tentative one because our calculation assumes a single unitary elasticity of substitution among all inputs and factors. Further calculations should be made concerning elasticities of substitution and their influence on measurements of the pattern of industrial protection.[19]

Finally, the data indicate, and it comes as no surprise, that tariffs tend to be higher on processed goods of interest to potential exporting *LDC*s than tariffs on other processed goods. No conclusion could be drawn concerning the size of protection of freight factors for the two groups of processed goods (except that the level is not insignificant for either grouping).

Thus, in spite of the gross level of aggregation, this study has been able to shed some light on the relative importance of the protection due to international transport costs vis-à-vis tariffs. The findings should not be embraced without reservation however. Better disaggregation would be much more informative as would focusing on a particular trade route or trading partner.

[19] Leith's paper is the major contribution on this matter.

REFERENCES

J. E. Anderson and S. Naya, "Substitution and Two Concepts of Effective Rate of Protection," *Amer. Econ. Rev.*, Sept. 1969, *59*, 607–12.

B. Balassa, "Tariff Protection in Industrial Nations and Its Effects on the Exports of Processed Goods from Developing Countries," *Can. J. Econ.*, Aug. 1968, *1*, 583–94.

———, "Tariff Protection in Industrial Countries, an Evaluation," *J. Polit. Econ.*, Dec. 1965, *73*, 573–94.

G. Basevi, "The United States Tariff Structure: Estimates of Effective Rates of Protection of United States Industries and Industrial Labor," *Rev. Econ. Statist.*, May 1966, *48*, 147–60.

W. M. Corden, "The Structure of a Tariff System and the Effective Protective Rate," *J. Polit. Econ.*, June 1966, *74*, 221–37.

J. E. Floyd, "The Overvaluation of the Dollar: A Note on the International Price Mechanism," *Amer. Econ. Rev.*, Mar. 1965, *55*, 95–107.

H. G. Johnson, "The Theory of Effective Protection and Preferences," *Economica*, May, 1969, *36*, 119–38.

———, "The Theory of Tariff Structure, with Special Reference to World Trade and Development," *Trade and Development*, Geneva 1965.

H. F. Karreman, *Methods for Improving World Transportation Accounts, Applied to 1950–1953*, New York Nat. Bur. Econ. Res., technical paper 13, 1961.

J. C. Leith, "Substitution and Supply Elasticities in Calculating the Effective Protective Rate," *Quart. J. Econ.*, Aug. 1968, *82*, 588–601.

C. Moneta, "The Estimation of Transportation Costs in International Trade," *J. Polit. Econ.*, Feb. 1959, *67*, 41–58.

W. G. Waters II, "The Protection Effect of International Transport Costs," unpublished doctoral dissertation, Univ. Wisc., June 1969.

NOTES

In accordance with the By-Laws of the American Economic Association, the President-elect has appointed a Nominating Committee for 1971, as follows: Kenneth Boulding, University of Colorado, Chairman; Michael Boskin, Stanford University; Hollis Chenery, International Bank for Reconstruction and Development and Harvard University; Michael Lovell, Wesleyan University; Alice Rivlin, The Brookings Institution; John H. Young, University of British Columbia and The Prices and Incomes Commission, Ottawa. Members of the Association are invited to send suggestions of nominees for various offices to the Chairman.

Omicron Delta Epsilon, The Honor Society in Economics, announces the 1970 winners of the Frank W. Taussig and Irving Fisher Awards. The Taussig Award was won by Miss Sharon Oster of Hofstra University with her entry "Are Black Incomes More Unequally Distributed?". The Fisher Award was won by Charles Nelson of the University of Chicago with his entry "The Term Structure of Interest Rates" which was his Ph.D. dissertation from the University of Wisconsin.

The entries for the third year of the competitions should be submitted to the Departmental Selection Committees by January 1, 1971. Winners to be judged by Selection Board consisting of Professors Kenneth J. Arrow, Kenneth E. Boulding, Milton Friedman, Paul Samuelson, Egon Neuberger (Editor). The Taussig Award consists of $100 and publication in *The American Economist*. The Fisher Award consists of $1,000 and publication by Basic Books, Inc.

Any undergraduate or recent graduate from a department with a chapter of *ODE* is eligible to enter the Taussig competition. Any graduate student or recent Ph.D. who is a member of *ODE* is eligible to enter the Fisher competition. For more information write to: Egon Neuberger, Editor, Department of Economics, State University of New York, Stony Brook, New York 11790.

New Journal: *The Review of Regional Studies*

First publication is spring 1971. The *Review* is a joint effort by the Southeastern Regional Science Association and the College of Business, Virginia Polytechnic Institute and State University. Papers are scheduled to appear from contributors in major fields to include geography, sociology, marketing, finance, urban studies, education, economics, and political science. For further information, write: Managing Editor, Stanley E. Boyle, Virginia Polytechnic Institute, College of Business, Blacksburg, Virginia 24061.

In 1971 the *Social Science Quarterly* will publish an issue on Urban Problems and Policies, and invites manuscripts dealing with these topics. Articles should be no longer than thirty typewritten pages (including footnotes and tables) and research notes of ten pages or less. Receipt of three copies of a manuscript usually facilitates an editorial decision within six weeks. For Style and format see recent issues of *SSQ* or write for style sheet. Manuscripts will be considered until June 15, 1971 and should be addressed to the Editor, *Social Science Quarterly*, The University of Texas, Austin, Texas, 78712.

The National Institute of Child Health and Human Development announces the initiation of a Population Research Centers Program to augment the current support mechanisms for population research. The program will be administered by the Center for Population Research which was established in August 1968, to undertake a long-term research effort in the social and behavioral sciences dealing with the causes and consequences of population growth, structure, and change, as well as to expand research efforts in reproductive biology with the goal of the development of a variety of new methods of fertility regulation.

The Population Research Centers Program is intended to encourage the creation and maintenance of environments favorable to research on population problems not readily investigated in other settings. It is anticipated that two to four Population Research Centers will be supported during the fiscal year 1971. Applications to be considered for this period must have been received by October 1, 1970. Thereafter, the deadlines will be February 1, June 1, and October 1.

Detailed information concerning the program and application forms may be obtained by writing: Dr. James F. O'Donnell, Program Director, Population Grants Branch, Center for Population Research, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20014.

Stanford University and the Inter University Committee on Urban Economics will sponsor a Summer Institute in Urban Economics for college and university teachers of economics, June 27-July 24, 1971. The program is supported by the National Science Foundation.

Twenty-five participants will be selected. The purposes of the Institute are to train the participants in the subject matter and methodology of urban economics, and to assist them to build courses of instruction in urban economics at their own institutions. For application forms and more information, write: Professor Henry M. Levin, *NSF* Institute in Urban Economics, Institute for Public Policy Analysis, Stanford University, Stanford, California 94305.

The department of economics of the State University of New York at Binghamton has been awarded a National Science Foundation development grant. The

purpose of the grant is to enable the department to develop a graduate program oriented around the theory, processes, and consequences of economic growth. The grant provides generously for graduate student financial assistance, as well as for new faculty additions. In order to implement the focus on policy problems arising from growth, the department will be emphasizing recruitment in such fields as urban economics, industrial organization and microeconomic theory.

For further information on either student assistantships or faculty positions write to Professor John La Tourette, Chairman, Department of Economics, State University of New York, Binghamton, New York, 13901.

The U.S. Office of Education's National Center for Educational Research and Development announces its fiscal year 1971 program of basic research in education. Proposals are invited from economists and researchers in other fields for projects which potentially will contribute to ultimate improvements in education. Research that has a strong theoretical orientation and shows promise of strengthening the scientific knowledge base is encouraged. Proposals are reviewed by panels of individuals with appropriate expertise from the academic and scientific community. The postmark deadline for the submission of proposals is February 13, 1971. Results of evaluations will be available approximately three months after the deadline. Guidelines for the submission of proposals may be obtained at the following address: Research Analysis and Allocation Staff, National Center for Educational Research and Development, U.S. Office of Education, 400 Maryland Avenue S.W., Washington, D.C. 20202.

Authors of papers for possible inclusion in the Medical Care Section Program of the next annual meeting of the American Public Health Association, October 11–15, 1971 in Minneapolis, Minnesota, may obtain standard abstract forms from: Mr. D. Brian Heller, Senior Director of Special Studies, Research and Development, Blue Cross Association, 840 North Lake Shore Drive, Chicago, Illinois 60611. Two types of papers will be considered: research reports and descriptions of programs or demonstrations. The deadline for submitting abstracts is April 15, 1971. Those whose papers are selected for the program will be notified in early June 1971.

The Surveys of Economic Opportunity are now available for use by interested researchers involved in questions concerning the causes and correlates of poverty in the United States. These two surveys were taken in 1966 and 1967 to augment the information regularly collected in the Current Population Surveys. In addition to a number of items common to both surveys, the *SEO* also provides information on other characteristics such as housing, marital history, training, assets and liabilities for a sample of about 30,000 households, which includes a supplementary sample (approximately 12,000) in areas with a large concentration of nonwhites. For more information contact: Data and Computation Center for the Social Sciences, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706.

The third Amos Tuck School Seminar on Problems of Regulation and Public Utilities, sponsored by the American Telephone and Telegraph Company, will be held during the first week of September, 1971, at Dartmouth College, Hanover, New Hampshire. The Seminar, which will be attended by thirty participants drawn from academic institutions throughout the country, consists of seven or eight sessions over a four-day period. The sessions, each of which will focus on a single paper, provide a forum for the presentation and discussion of new ideas and innovative work in analyzing regulated firms, regulatory processes, and problems of public welfare.

Those chosen to present papers will receive up to $1,200 in support of research and writing; travel and living expenses during the Seminar are also covered. Particular attention will be given to papers submitted by scholars in the beginning stages of their careers or who have only recently turned their attention to public utility problems. For further information, write: Professor Richard S. Bower or Professor Willard T. Carleton, Co-Directors of the Seminar, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

Frederic J. Fleron, Jr. has recently become editor of the *Newsletter on Comparative Studies of Communism* sponsored by the Planning Group on Comparative Communist Studies of the American Council of Learned Societies. The *Newsletter* will carry announcements of conferences, research projects, and archives. It will also carry review essays on books and articles, commentaries and debates, and will serve as a forum for the exchange of views on issues in the comparative study of Communism. Persons wishing to have their names added to the mailing list or desiring to make contributions to the *Newsletter* or "Occasional Papers" should write to Professor Frederic J. Fleron, Jr., Editor, *Newsletter on Comparative Studies of Communism*, Department of Political Science, State University of New York at Buffalo, 4238 Ridge Lea Road, Buffalo, N.Y. 14226.

Deaths

Seymour Morosoff, Delmar, New York, May 9, 1970.

Harold C. Taylor, director, The W. E. Upjohn In-

stitute for Employment Research, Kalamazoo, Michigan, May 6, 1970.

Jacob Viner, professor of economics and international finance, Princeton University, September 12, 1970.

Retirements

Roy Blough, professor of banking and international finance, Columbia University, June 30, 1970.

John V. Conner, professor of economics, Loyola University, New Orleans, June 1970.

Oskar Morgenstern, Princeton University, June 1970.

Arthur R. Olsen, professor of economics, Western Illinois University, August 31, 1970.

Herman F. Otte, professor of economic geography, Graduate School of Business, Columbia University, June 30, 1970.

Joseph H. Taggart, executive dean, School of Business, New York University, July 1970.

Demetrios Theodore, department of economics, Cleveland State University, Sept. 1970.

Visiting Foreign Scholars

Robert Boland, Graduate School of Business, University of Cape Town, South Africa: visiting professor, Graduate School of Business, Columbia University, June 1, 1970.

Frans Alting von Geusau, John F. Kennedy Institute, Tilburg: visiting professor, Drew University.

Maurice McManus, University of Birmingham: visiting professor of economics, University of Hawaii, 1970–71.

Gunnar Myrdal, Institute for International Economic Studies, Stockholm: visiting scholar, University of Georgia and Atlanta University, fall 1970.

Francis Seton, Nuffield College, Oxford: visiting professor of economics, University of Waterloo, fall term.

Brinley Thomas, University College, Cardiff; Cardiff, Wales: visiting professor of economics, Brown University, spring 1971.

Adolf Weber, Christian-Albrechts-Universität, Holzkoppelweg, West Germany: visiting professor of agricultural economics, University of Minnesota.

Jean Waelbroeck, University of Brussels: visiting professor, Drew University.

Promotions

Michael Adler: associate professor, Graduate School of Business, Columbia University, July 1, 1970.

Edward I. Altman: associate professor of finance, New York University.

Wallace Atherton: professor of economics, California State College, Long Beach.

James F. Becker: professor of economics, School of Commerce, New York University.

Ben W. Bolch: associate professor of economics, Vanderbilt University, Sept. 1970.

George Budzeika: special assistant, banking studies department, Federal Reserve Bank of New York.

John C. Burton: professor, Graduate School of Business, Columbia University, July 1, 1970.

Kathleen Q. Camin: associate professor of economics, Wichita State University.

Virgil L. Christian, Jr.: professor of economics, University of Kentucky.

Dale C. Dahl: professor of agricultural economics, University of Minnesota.

Elizabeth Durbin: assistant professor of economics, School of Commerce, New York University.

Eldon J. Dvorak: professor of economics, California State College, Long Beach.

John Farley: professor, Graduate School of Business, Columbia University, July 1, 1970.

T. Aldrich Finegan: professor of economics, Vanderbilt University, Sept. 1970.

James Fortson: associate professor of finance and forestry, University of Georgia.

Richard E. Gift: associate professor of economics, University of Kentucky.

Robert Holbrook: associate professor of economics, University of Michigan.

David Holdsworth: special assistant, banking studies department, Federal Reserve Bank of New York.

James P. Houck, Jr.: professor of agricultural economics, University of Minnesota.

John G. Hutchinson: professor, Graduate School of Business, Columbia University, July 1, 1970.

Saul Hymans: professor of economics, University of Michigan.

George Johnson: associate professor of economics, University of Michigan.

Louis H. Jordan: professor, Graduate School of Business, Columbia University, July 1, 1970.

Warren J. Keegan: associate professor, Graduate School of Business, Columbia University, July 1, 1970.

W. Michael Keenan: associate professor of finance, New York University.

Sun K. Kim: associate professor, department of economics and statistics, California State College, Los Angeles, Sept. 1970.

Charles Lamphear: associate professor of economics, University of Nebraska.

Gurcharan S. Laumas: professor of economics, Kent State University.

Nathaniel Leff: associate professor, Graduate School of Business, Columbia University, July 1, 1970.

Jimmie R. Monhollon: senior vice president in charge, Federal Reserve Bank of Richmond, Charlotte.

Randall I. Mount: associate professor of economics, Kent State University.

George Parker: associate professor, Graduate School of Business, Columbia University, July 1, 1970.

Mark V. Pauly: associate professor of economics, Northwestern University.

Jerry Petr: associate professor of economics, University of Nebraska.

Edward J. Powers: associate professor of economics, Northern Michigan University.

Peter A. Prosper, Jr.: associate professor, department of economics, Union College.

Malcolm J. Purvis: associate professor of agricultural economics, University of Minnesota.

Robert A. Schwartz: associate professor of economics and finance, New York University.

Harold T. Shapiro: professor of economics, University of Michigan.

J. Edward Smith, Jr.: assistant professor, department of economics, Union College.

Robert W. Snyder: associate professor and extension land economist, agricultural economics and extension service, University of Minnesota.

George J. Staller: professor of economics, Cornell University.

G. Frederick Starner: assistant professor of economics, Drew University.

William T. Terrell: associate professor of economics, Wichita State University.

Kenneth H. Thomas: associate professor and extension economist, farm management, agricultural economics and extension service, University of Minnesota.

Henry Thomassen: professor of economics, University of Nebraska.

John J. Waelti: associate professor and extension economist, agricultural economics and extension service, University of Minnesota.

E. Kirby Warren: professor, Graduate School of Business, Columbia University, July 1, 1970.

Samuel C. Webb: associate professor of economics, Wichita State University.

Alan R. Winger: professor of economics, University of Kentucky.

Henry W. Woudenberg: associate professor of economics, Kent State University.

Administrative Appointments

Joseph C. Blumel: vice president for academic affairs, Portland State University.

Kenneth T. Cann: chairman, department of economics, Western Kentucky University, Sept. 1970.

Rudolf Coper: assistant dean, College of Business Administration, Loyola University, Ned Orleans, April 1970.

E. Gerald Corrigan: chief, domestic research division, Federal Reserve Bank of New York.

Frederic N. Firestone: chairman, department of economics, Illinois State University.

Robert G. Hawkins: professor, chairman, department of finance, Graduate School of Business Administration, New York University.

Arthur P. Hurter, Jr.: professor, chairman, department of industrial engineering and management sciences, Northwestern University, Sept. 1970.

John B. Lansing: chairman, department of economics, University of Michigan.

Leonard Lapidus: assistant vice president, Federal Reserve Bank of New York.

T. C. Liu: chairman, department of economics, Cornell University.

H. Michael Mann: chairman, department of economics, Boston College, July 1, 1970.

Jürg Niehans: chairman, department of political economy, Johns Hopkins University.

John A. Orr, Florida Atlantic University: chairman, department of economics, Chico State College.

Harold Petersen: vice chairman, department of economics, Boston College, July 1, 1970.

Scott E. Pardee: assistant vice president, Federal Reserve Bank of New York.

Vernon W. Ruttan: director, Economic Development Center, University of Minnesota.

Edward K. Smith: vice president, research services and planning, National Bureau of Economic Research.

Shanti S. Tangri, Duke University: professor, chairman, department of economics, Livingston College, Rutgers University.

Felino J. Valiente: professor, chairman, department of accounting, College of Business Administration, Loyola University.

New Appointments

Daniel J. Ableman, Texas A&M University: assistant professor of economics, University of Georgia, Sept. 1970.

C. D. Acland: associate professor, department of economics, Carleton University.

Jill Adler: economist, domestic research division, Federal Reserve Bank of New York.

Armen Alchian, University of California, Los Angeles: visiting professor of economics, University of Hawaii, fall 1970.

Margaret Alumkal: associate professor of management, College of Business Administration, Loyola University, New Orleans, 1970–71.

Ross Azevedo: instructor, department of economics, State University of New York, Brockport.

Richard Bartel: economist, domestic research division, Federal Reserve Bank of New York.

Clarence D. Billings: assistant professor of finance and institute of law and government, University of Georgia, fall 1970.

Erwin A. Blackstone, Dartmouth College: assistant professor of economics, Cornell University.

Frederick Breimyer: economist, domestic research division, Federal Reserve Bank of New York.

John A. Broaddus, Indiana University: economist, Federal Reserve Bank of Richmond.

Cynthia Brown: instructor, economics department, Barnard College, 1970–71.

Galen D. Burghardt, Jr.: assistant professor, department of economics, University of Massachusetts.

H. Stuart Burness: assistant professor of economics, University of Kentucky.

Eveline Burns: visiting professor, department of economics, Barnard College, fall 1970.

John Chu: economist, balance of payments division, Federal Reserve Bank of New York.

James F. Cook, Virginia Polytechnic Institute: assistant professor of economics, Winthrop College.

James C. Cox: assistant professor, department of economics, University of Massachusetts.

Elizabeth Crowell, Indiana University: assistant professor of economics, Northern Michigan University.

Suzanne Cutler: economist, banking studies department, Federal Reserve Bank of New York.

Philip H. Davidson, University of Illinois: economist, Federal Reserve Bank of Richmond.

Alan Deardorff, Cornell University: assistant professor, department of economics, University of Michigan.

Peter d'Errico: assistant professor, department of general business and finance, University of Massachusetts, Sept. 1, 1970.

Albert de Prince: economist, market statistics division, Federal Reserve Bank of New York.

Stanley Diller: assistant professor, Graduate School of Business, Columbia University, July 1, 1970.

Thomas E. Duston: assistant professor of economics, University of Massachusetts.

James P. Egan, La Moyne College: economics department, Wisconsin State University.

Randy Elkin: assistant professor of economics, Illinois State University.

Stuart M. Feder: economist, foreign research division, Federal Reserve Bank of New York.

Glenn W. Fisher, University of Illinois: professor of urban affairs, Wichita State University.

Moheb Ghali: assistant professor of economics, University of Hawaii.

E. Kenneth Grant: assistant professor, department of economics, University of Guelph, July 1970.

John K. Green: assistant professor of economics, University of the South, Sept. 1970.

Stephen Guisinger: assistant professor of economics, Southern Methodist University, fall 1970.

Kanji Haitani, Southern Illinois University: associate professor of economics, State University of New York, Fredonia.

Judd M. Hammack, University of Washington: assistant professor, department of economics and statistics, California State College, Los Angeles.

Randall B. Hayden, Indiana University: professor of money and banking, Wichita State University.

H. Robert Heller, University of California (Los Angeles): visiting associate professor of economics, University of Hawaii, fall 1970.

Walter Hettich: associate professor, department of economics, Carleton University.

Lawrence J. Hexter, University of Maryland: associate professor of finance, Kent State University.

Seiko Higa: assistant professor, department of economics, State University of New York, Brockport.

John C. Holmstrom: assistant professor of economics, University of Hawaii.

Thomas M. Humphrey, St. Andrews College: economist, Federal Reserve Bank of Richmond.

Thomas Ireland: assistant professor of economics, Illinois State University.

Eugene J. Jaffe, fiscal division, U.S. Marine Corps: U.S. Army procurement center, Fort Lee.

Ethan Katch: visiting assistant professor, department of general business and finance, University of Massachusetts, Sept. 1, 1970.

Donald W. Katzner, University of Pennsylvania: associate professor of economics, University of Waterloo.

Patricia Keough: economist, domestic research division, Federal Reserve Bank of New York.

Richard E. Kihlstrom, University of Minnesota: visiting assistant professor, department of economics, University of Massachusetts.

Glenn Kirby, University of Nebraska at Omaha: assistant professor of economics, Millikin University, fall 1970.

Patricia Kuwayama: economist, foreign research division, Federal Reserve Bank of New York.

Leonard L. Lederman, Battelle Memorial Institute: deputy head, Office of Economic and Manpower Studies, National Science Foundation.

Ronald Lee, Harvard University: assistant professor, department of economics, University of Michigan.

George Lermer, Sir George Williams University: associate professor of economics, University of Waterloo.

Fred J. Levin: economist, domestic research division, Federal Reserve Bank of New York.

Shu-jan Liang: assistant professor of economics, College of Business Administration, Loyola University.

Richard G. Lipsey: professor, department of economics, Queen's University.

Isaiah A. Litvak: professor, department of economics, Carleton University.

Donald L. Losman: associate professor of economics, College of Business Administration, Loyola University.

Albert Lowey-Ball: instructor in economics, Drew University.

James R. McCabe, Western Illinois University: economist, Federal Reserve Bank of Richmond.

John J. McCall, University of California, Irvine: professor, department of economics, University of California, Los Angeles.

Robert T. McKinnell: associate professor, department of economics, Carleton University.

James Mak: assistant professor of economics, University of Hawaii.

C. J. Maule: associate professor, department of economics, Carleton University.

Edward F. Meeker, University of Washington: assistant professor of economics, University of Georgia, Sept. 1970.

Ann-Marie Meulendyke: economist, domestic research division, Federal Reserve Bank of New York.

John R. Morris, Jr.: assistant professor of economics, University of Colorado, Sept. 1970.

Lynn Muchmore, Grinnell College: assistant professor of economics, Cornell University.

Thomas J. Muench, University of Minnesota: visit-

ing associate professor, department of economics, University of Massachusetts.

Frank Muller, University of Bochum, West Germany: assistant professor of economics, University of Waterloo.

Eliot S. Orton, Cornell University: assistant professor, department of economics, New Mexico State University, Sept. 1970.

Jan Overbeek: assistant professor of economics, University of Hawaii.

Joseph E. Peno: instructor, department of economics, Union College.

Elliott Platt: economist, domestic research division, Federal Reserve Bank of New York.

John L. Pratschke, The Economic and Social Research Institute, Dublin: assistant professor of economics, University of Waterloo.

David Pritchett: economist, domestic research division, Federal Reserve Bank of New York.

K. P. Prodromidis: assistant professor department of economics, Carleton University.

Joseph Rabianski: assistant professor, department of economics, State University of New York, Brockport.

Wolfhard Ramm: assistant professor of economics, University of California, San Diego.

Richard D. Raymond, West Virginia University: associate professor of economics, Kent State University.

Michael Rieber, University of Missouri: professor of economics, Kent State University.

Petronio L. Riós, Institute Forestal Casilla 3085, Santiago: assistant professor, department of economics and statistics, California State College, Los Angeles, Mar. 1970.

Jack L. Rutner: assistant professor of economics, Illinois State University.

Leonard G. Sahling: economist, domestic research division, Federal Reserve Bank of New York.

Warren C. Sanderson: research staff associate, National Bureau of Economic Research.

Fredricka P. Santos: instructor, economics department, Barnard College.

Robert J. Saunders, West Virginia University: associate professor of economics, Kent State University.

Gary Saxonhouse, Yale University: assistant professor, department of economics, University of Michigan.

Albert A. Schepanski: assistant professor, department of economics, Carleton University.

Richard Schmalensee: assistant professor of economics, University of California, San Diego.

Bernard Schneider: economist, balance of payments division, Federal Reserve Bank of New York.

Lawrence E. Schwartz, Mitre Corporation: associate professor, department of economics, University of Utah.

Andrew Seidel: economist, market statistics division, Federal Reserve Bank of New York.

Jacobus T. Severiens: temporary assistant professor of banking and finance, University of Georgia, fall 1970.

Wayne W. Sharp: agricultural attache, U.S. Embassy, Guatemala City, Guatemala.

Richard A. Shick: temporary assistant professor of banking and finance, University of Georgia, fall 1970.

Paul Y. Shin, University of Hawaii: department of economics and statistics, California State College, Los Angeles, Sept. 1970.

Donald Shoup, Yale University: assistant professor, department of economics, University of Michigan.

David B. Smith, University of Nebraska: assistant professor, department of economics, New Mexico State University, Sept. 1970.

Donald M. Smith: assistant professor, department of economics, Southern Methodist University, fall 1970.

Thomas E. Snider, Virginia Polytechnic Institute: economist, Federal Reserve Bank of Richmond.

Hugo Sonnenschein, University of Minnesota: professor, department of economics, University of Massachusetts.

Thomas Sowell, Brandeis University: associate professor, University of California, Los angeles.

Gary Stern: economist, domestic research division, Federal Reserve Bank of New York.

David K. Stout: research staff associate, National Bureau of Economic Research.

Rao Tatikonda: assistant professor of management, College of Business Administration, Loyola University, New Orleans.

David Vanderford: economist, market statistics division, Federal Reserve Bank of New York.

Peter Van der Spek: economist, foreign research division, Federal Reserve Bank of New York.

Ralph von Gersdorff: visiting professor of economics, St. Francis Xavier University.

Henry T. Wan, Univeristy of California, Davis: professor of economics, Cornell University.

Peter B. Watson: assistant professor of economics and transportation, Northwestern University.

Edwin G. West: professor, department of economics, Carleton University.

Clifford E. Wheeler, Westminister College: assistant professor of economics, Wichita State University.

Harold F. Williamson: visiting scholar, Eleutherian Mills Hagley Foundation Library, fall 1970.

Robert J. Willis: research fellow, National Bureau of Economic Research.

T. Hinds Wilson, University of North Carolina: visiting research associate, Federal Reserve Bank of Richmond.

Gerald Zeisel: assistant professor of accounting, school of business administration, University of Massachusetts, Sept. 1, 1970.

Michael Zubkoff: assistant professor of economics, Vanderbilt University.

Leaves for Special Appointments

John Beckett, Whittemore School of Business and Economics, University of New Hampshire: visiting professor of management, Graduate School of Busi-

ness, Columbia University, summer term, June 1, 1970.

Robert W. Beckstead, New Mexico State University: Department of the Army, Safeguard System Evaluation Agency, White Sands Missile Range, New Mexico.

W. Keith Bryant, University of Minnesota: Poverty Center, University of Wisconsin.

William L. Campfield, Office of Policy and Special Studies, U.S. General Accounting Office, Washington, D.C.: visiting professor of accounting, Graduate School of Business, Columbia University, autumn term, Oct 1, 1970.

Charles H. Cuykendall, University of Minnesota: Cornell University.

Joseph M. Finger, University of Georgia: U.S. Tariff Commission, Washington, 1970–71.

Patrick J. Gormely, Kansas State University: U.S. Agency for Internation Development, Lagos, Nigeria.

William E. Griffiths, University of Illinois: research associate in agricultural economics, University of Minnesota.

Michael J. Hamburger, Federal Reserve Bank of New York: Bank of England.

Derek T. Healy, University of Adelaide, South Australia: research fellow, Institute for Industrial Economic Studies, Stockholm, Sweden.

David H. Kopf, Federal Reserve Bank of New York: economic advisor and manager of research department, Bank of Tanzania.

Joseph Krislov, University of Kentucky: Fulbright Lectureship at Trinity College, Dublin, 1970–71.

James W. McKie, Vanderbilt University: Senior Fellow, The Brookings Institution.

Frank J. Macchiarola, Bernard M. Baruch College: assistant professor of law, Graduate School of Business, Columbia University, autumn and spring terms, July 1, 1970.

Walter P. Page, University of Georgia: University of Kansas, 1970–71.

Gian S. Sahota, Vanderbilt University: visiting scholar, Harvard University, fall 1970; Simon Senior Fellow, University of Manchester, spring and summer 1971.

Donald Schilling, University of Missouri-Columbia: visiting associate professor, department of economics, Southern Methodist University, 1970–71.

Hans Schollhammer, University of California, Los Angeles: visiting assistant professor, Graduate School of Business, Columbia University, autumn term, Oct. 1, 1970.

John A. Speicher, Michigan State University: visiting associate professor of agricultural economics, University of Minnesota.

Jeremy L. Wiesen, Wharton School: visiting assistant professor of business law, Graduate School of Business, Columbia University, spring and summer terms, Feb. 1, 1970.

Rozo Yamamura, Boston College: visiting professor, University of Washington.

Resignations

Loren E. Casement, University of Nebraska: Westminister College, Sept. 1970.

Donald E. Farrar, Graduate School of Business, Columbia University, June 30, 1970.

Lyndell W. Fitzgerald, University of Minnesota: Food and Agriculture Organization of the United Nations, Rome, Italy, June 1, 1970.

Fred H. Goldner, Graduate School of Business, Columbia University, June 30, 1970.

Robert L. Greene, University of Georgia, June 1970.

Philip W. Jeffress, Western Kentucky University, Aug. 1970.

Alain C. Jones, College of Business Administration, Loyola University, New Orleans.

Alfred L. Kahl, Jr., University of Georgia, June 1970.

Robert Keller, College of Business Administration, Loyola University, New Orleans, May 1970.

L. Ledohowski, Carleton University.

Peter B. Lund, Vanderbilt University: Sacramento State College.

Dean Morse, Graduate School of Business, Columbia University, June 30, 1970.

R. Richard Ritti, Graduate School of Business, Columbia University, June 30, 1970.

W. R. Scott, Carleton University: Queen's University.

Russell Taussig, Graduate School of Business, Columbia University, June 30, 1970.

George van Furstenberg, Cornell University: University of Indiana.

Wynnelle Wilson, Federal Reserve Bank of Richmond: Florida State University.

Jack Zwick, Graduate School of Business, Columbia University, June 30, 1970.

Miscellaneous

Lawrence A. Mayer: board of editors, Fortune Magazine.

SIXTY-SEVENTH LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1970. Abstracts of many of the dissertations are supplied.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

ASSIBI ABUDU, Ph.D. California (Los Angeles). Establishing gold mining rights in California, 1848–1853.

Conventional economic analysis inadequately considers the role of property rights in determining the value of resurces. California provided a case where valuable resources, without definite property rights, were subjected to the law of capture. Wealth maximizing actions by miners are examined when the ratio of the number of miners to value of land was lowest, and when this ratio declined. The careful definition of miners' rights vis-á-vis those of non-miners was also examined.

HIDEYUKI ADACHI, Ph.D. Rochester 1970. The vintage approach to the theory of capital and growth.

This research investigates various properties of a putty-clay model. The physical durability of investment as well as its economic lifetime is treated as one of the endogenous variables in the model. The choice of technique (represented by various combinations of labor intensity of and durability of investment) of a firm, the determination of production techniques and factor prices in a two-sector economy with putty-clay technology are discussed.

EPHRAIM ASHER, Ph.D. Rochester 1970. Relative productivity, factor intensity and technology in the manufacturing sectors of the United States and the United Kingdom during the nineteenth century.

Theoretical and empirical examination of Habakkuk's contention of biased technological change in American and British industries. Under the assumption of identical technologies, the examination of relative factor prices and surrogates for factor intensity show that American textiles and iron industries were more capital intensive than their British counterparts. Under an alternative assumption of different technologies given by CES production functions, labor-saving bias and capital-saving bias are measured in American and British textiles, respectively.

JAMES L. BARR, Ph. D. Stanford 1970. Empirical implications of a vintage production model.

JOHN P. BARRADOS, Ph.D. Columbia 1970. Studies in the dichotomy in the theory of price.

The dissertation shows, contrary to the findings of Lange and Patinkin, that the classical real and monetary analyses can be integrated in a consistent manner. The classical economists developed self-contained real and monetary analyses. These can be integrated, as shown in the dissertation, only after a logical, economic relationship between them is introduced such as, for example, making the monetary commodities in the real sector.

RAYMOND C. BATTALIO, Ph.D. Purdue 1970. An investigation of a class of technological theories of production.

RALPH E. BEALS, Ph.D. Massachusetts Institute of Technology 1969. Utility-of-wealth functions for speculators in commodity futures.

BRIAN L. BENTICK, Ph.D. Yale 1969. Saving, investment, and the land market: Victoria, 1872–93.

HENRY W. BOWMAN, Ph.D. Syracuse 1969. Price behavior in underinformed markets.

DAGOBERT L. BRITO, Ph.D. Rice 1970. On the limits of economic control.

There has been a controversy whether, given the lags and uncertainty in the economy, the central authorities should react to short-run fluctuations of the system or use a constant control (e.g., constant rate of growth of money supply). This dissertation investigates the problem for a system whose behavior is described by a system of differential equations. Loosely, the results are that under certain conditions there exists a region in state space where the constant controls are "superior."

MEREDITH N. BROWNE, Ph.D. Texas (Austin) 1969. The corporation in American economic thought, 1830–1930.

The evolution of the private business corporation in the United States elicited numerous and varied reactions from economic analysts writing during the period 1830–1930. In this study these responses are organized and evaluated in an effort to trace the intellectual origins of contemporary understanding of the corporation as a leading economic institution. The manner in which and the extent to which economic thought adapted to the maturation of the modern corporation constitute the central problem under study.

CARLTON J. BRUCE, Ph.D. Rochester 1969. The role of expectations in resource allocation and growth.

This project analyzes the behavior of a neoclassical growth model in which both capital heterogeneity and price expectations are explicitly taken into account. In the descriptive problem, it examines the conditions, particularly the factor-intensity conditions which guarantee uniqueness of equilibrium and those which

guarantee dynamic stability under perfect foresight. It also examines the corresponding optimal problem, clarifies the relationship between the two problems, and investigates the possibility of solution to the inverse optimal problem.

ROBERT C. BUSHNELL, Ph.D. Princeton 1970. The linear assumption in the use of the linear programming model in economics.

The assumption of linearity in the absence of better knowledge is rejected and the consequences of the linear assumption explored. It is suggested that the linear programming model is not generally applicable as a theoretical model but useful as an approximative computational device when the feasible region is restricted. In characterizing the term "approximately linear," procedures are described to calculate limits to the range of the objective function for any quantification of objective function variation.

WAN-FU CHI, Ph.D. Southern Illinois 1969. Multiple goals in a firm.

A firm pursuing multiple goals must rank them by priority because efforts to optimize them may conflict with one another. Using the concept of distance delineated by six axioms—one of which is that the minimum positive distance is one—the optimum ranking of all environmental achievable rankings is obtained in one of two ways: minimization of the distance or minimization of the square of the distance over the range of the ranking function. The method is extended to a sequence of rankings on a line.

JAMES E. CIECKA, Ph.D. Purdue 1970. An expedient empirical interpretation and tests of theories of consumer demand.

RAYMOND E. DACEY, Ph.D. Purdue 1970. Aspects of two methodological issues in quantitative economic history.

The thesis explicates two methodological concepts of quantitative economic history: the data problem and counterfactual argument. The thesis serves two purposes in addition to presenting the desired explications. The explications are presented in a completely formal language; an advanced form of general predicate logic with identity and functors. Second, a precise formulation is made of those aspects of the methodological and philosophical foundations of quantitative economic history that are required for the explications. Included in those foundations are a theory of measurement and a quantitative inductive logic.

DONNA K. DIAL, Ph.D. Florida State 1969. Economic analysis of agrarian movement and social protest literature in the United States 1865–96.

This work deals with the protest literature of the late nineteenth century. Basically it examines the protest literature in light of the dominant economic issues of the period and contrasts the treatment of economic protest themes by such writers as Howell, Garland, and Mark Twain, with those of the so-called heterodox economists, Ely, Walker, Ross, and Veblen.

WALTER E. DIEWERT, Ph.D. California (Berkeley) 1969. Functional form in the theory of production and consumer demand.

DAVID P. DOANE, Ph.D. Purdue 1969. Regional structure the cotton textile industry, 1880–1900.

A formal explanatory model is adduced to explain the shift in textile manufacturing from New England to the Southeast after 1880. Empirical interpretations are made of the arguments of the model, including investigations of regional product mix, technical efficiency, production costs by output class, and simulation of capital and transport costs for firms in various locations. Reactions to changing 1880–1900 market conditions are predicted for firms in the hypothesized decision environment.

RONALD W. FILANTE, Ph.D. Purdue 1969. An inquiry into the sources of surplus revenues on the Erie Canal, 1825–60.

Output, population, freight and passenger traffic, and revenue data for the entire antebellum economy of the North Central and Middle Atlantic states are incorporated into intra- and interregional models of production and trade. This procedure permits the identification and quantification of those factors responsible for the Erie Canal's preeminence among other canals built during the period, as well as its ability to operate successfully under conditions of competition with the emerging rail systems of the late antebellum years.

BENITO FLORES-SANDOVAL, Ph.D. Houston 1970. Optimum seeking methods and examples of its summary use in economics.

Utilizing a special quadratic function that is used for interpolation, a method was devised that can be used for the optimization of unrestricted non-linear functions. Because of its design, the method will at least yield local or near-local optimum points. The procedure was computer programmed and utilized to solve problems in economics. Comparison was made to other computational methods with favorable results.

GERALD E. FLUECKIGER, Ph.D. Purdue 1970. The structure and behavior of technological change in the iron and steel industry: 1700–1899.

Production is characterized in terms of a cybernetic system. Technical change is defined as the appearance or removal of names (products, factors, and machines) from lists (catalogues, payrolls, and capital inventories). Measures of technological change for the iron and steel industry are developed by using a chronology of innovations. Education and training are defined in terms of a cybernetic system. Some theorems about the specialization and division of labor are presented and related to the notions of education and training.

LOUIS C. GASPER, Ph.D. Duke 1969. Organized crime: an economic analysis.

MARWAN M. GHANDOUR, Ph.D. Illinois (Urbana) 1969. An optimal control model of capital accumulation and development following a Soviet strategy.

PHILIP M. GINSBERG, Ph.D. Purdue 1969. Consumption theory and variable preferences.

DANIEL A. GRAHAM, Ph.D. Duke 1969. Toward a dynamic theory of the firm.

JERRY R. GREEN, Ph.D. Rochester 1970. Some aspects of the use of the core as a solution concept in economic theory.

This thesis concentrates on two problems concerning the core of an economic system. For which economies is it true that at every point in the core identical individuals attain the same level of satisfaction? Let there be r_t identical individuals in the t^{th} type. It is shown that if g.c.d. $[r_t] \neq 1$ the above property holds. Further, for almost all economies failing to satisfy this condition, this result is false. The second problem is that of stability. Two adjustment mechanisms are treated. In each case it is supposed that the next proposal is selected stochastically from the set of all possible blocking proposals. It is shown that the probability of reaching the core approaches 1 with time.

ERIK F. HAITES, Ph.D. Purdue 1969. Ohio and Mississippi river transportation, 1810–60.

The study was confined to the commercial river craft—flatboats, keelboats, and steamboats—operating in the Louisville-New Orleans trade. The costs and revenues associated with the operation of these river craft were estimated for each decade. They confirmed contemporary reports of the early profitability of steamboats, steamboats driving the keelboats from this trade during the early 1820's, the unprofitability of steamboat operation after 1830, the emergence of a back-haulage problem for steamboats during the 1840's, and the decline of flatboating during the late 1840's and early 1850's.

JOHN F. HANIESKI, Ph.D. Purdue 1970. An explanatory model of technologically new products.

GEORGE A. HAY, Ph.D. Northwestern 1969. Production, price, and inventory theory: an integrated model of firm behavior.

By making the assumption that firms attempt to maximize the expected value of profits, we are able to derive empirically testable hypotheses regarding the principle determinants of firms' decisions on production, price, and finished goods inventory. Regressions are performed on data for several 2-digit SIC groups and the results are found generally to conform to the predictions of the model.

ROBERT F. HEBERT, Ph.D. Louisiana State 1970. A critical evaluation of Emile Cheysson's contributions to economic analysis.

Despite recognition by Leon Walras, Maffeo Pantaleoni, and Wilhelm Launhardt, the economics of the nineteenth century engineer-economist, Emile Cheysson, has received almost no attention by twentieth century economists. This dissertation, for which several obscure French sources were translated into English for the first time, concludes that Cheysson made original and important contributions to econometrics and managerical economics, particularly in matters of statistical revenue and cost functions, spatial economics, investment decisions, and product variation.

JOSEPH A. HESS, Ph.D. Northwestern 1970. Noncapital programming: a theory of financial interdependencies.

The firm is considered a system consisting of a trilogy of subsystems. One of the subsystems (noncapital) is defined as the short-run decision relating to the optimal allocation of scarce funds into finished goods inventory, accounts receivable, and advertising to induce the maximum outward shift in the demand function. Using mathematical programming techniques the interactions between the subsystems and between the decision variables are integrated into the model. Actual areas of decision making application are developed.

RICHARD S. HIGGINS, Ph.D. Virginia 1969. Economic theory and bureaucracy.

CHARLES F. HOLT, Ph.D. Purdue 1970. The role of state government in the nineteenth century American economy, 1820–1902: A quantitative study.

The study investigates long-term trends in state government involvement in the nineteenth century American economy as evidenced by categorized state Government money-flow transactions. Overlapping decade average estimates of the level and composition of aggregate state government expenditures and receipts for the period 1820–1902 are derived from the serial reports of state auditors and treasurers. The level of state expenditures maintained a stable relationship to national product, ranging between 0.5 and 1.1 percent. A revulsion against internal improvements is found to have occurred in the period 1840–54, marking a permanent shift in state development policy from transportation to education.

OSWALD HONKALEHTO, Ph.D. Massachusetts Institute of Technology. A theoretical study of technical substitution possibilities under joint production.

The range of output substitutability is defined and treated formally as a categorical variable with three possible values: zero, limited, and unlimited. Its pattern of variation is deduced for a number of theoretical variants of a basic pair of joint-production models—models that are fashioned from familiar technological and input-availability constraints. Considerable attention is then given to the input-output-substitution implications of these patterns of output substitutability.

DEAN T. JAMISON, Ph.D. Harvard 1970. Studies in individual choice behavior.

This thesis comprises a number of distinct essays

linked by a common theme—they all deal with one aspect or another of the theory of individual choice behavior. Topics covered include choices involving time; the psychological literature on time choices; an axiomatic treatment of choice involving time but not uncertainty; an initial attempt to allow for uncertainty; information and choice; a normative study; a range of theories designed to be empirically valid; and empirical studies relevant to the theoretical developments in earlier sections.

TETSUNORI KOIZUMI, Ph.D. Brown 1970. An analysis of factor substitution in the multi-factor production process.

The thesis gives a thorough examination of the concept of the elasticity of substitution in the production process utilizing more than two factor inputs. The usage of the concept is expanded to include applications in the theories of derived demand, production functions, and growth. Complementarity and limiting factors are considered in static distribution theory as well as in dynamic growth models.

DONALD F. KRIER, Ph.D. Boston College 1970. Population movements in England, 1650–1812: A family-reconstitution study of three eighteenth century Lancashire parishes.

A family-reconstitution study by cohort analysis of three Lancashire parishes was undertaken to determine the relative importance of the birth rate and the death rate in explaining population change in eighteenth century England. Such variables as age at first marriage of females and males, the average completed family size, and the age at death are measured and correlated with various occupational groups as recorded in the parish registers.

BEN E. LADEN, Ph.D. Johns Hopkins 1969. Profit maximization, average cost pricing and the price equation for U.S. manufacturing, 1954–66.

This study analyzes the empirical relationship between price of manufacturing output and cost and demand factors. Analysis of the process by which price changes occur is the primary purpose. Theoretical models are the basis for the empirical work. Three models—perfect competition, monopoly, and average cost pricing—are considered. The findings are inconsistent with results in the recent literature. In particular, no direct increase in price is implied by an increase in aggregate demand.

LAWRECE J-Y. LAU, Ph.D. California (Berkeley) 1969. Duality and utility structure.

RICHARD J. LURITO, Ph.D. Georgetown 1969. Returns to scale: theoretical aspects, estimation model, and simulation results.

The dissertation analyzes and tests an alternative approach to the problem of empirical scale elasticity estimation. While the cost curve and production function estimation approaches exhibit problems in obtaining unbiased scale elasticity estimates, their key problem lies in making required assumptions about the production function form. The proposed alternative approach requires no such assumptions; it not only permits unbiased estimates of scale elasticity to be made, but also determines the presence or absence of homogeneity and linear expansion paths.

DONALD N. MCCLOSKY, Ph.D. Harvard 1970. Economic maturity and entrepreneurial decline: British iron and steel, 1870–1913.

Writers on the British economy of the nineteenth century conclude that British entrepreneurs failed. The chief support of this view is the allegedly bad performance of the iron and steel industry, particularly its slower growth relative to Germany and the United States. Evidence on the growth of home demand, on the adoption of the basic production process, and on the growth of productivity indicate that this commonly accepted conclusion must be abandoned.

WILLIAM T. MCGRATH, Ph.D. Southern California 1969. On the theory of value and market syndicalism.

This study formulates a theory of an operationally efficient socialist economic system based on cooperative syndicalist firms. All production enterprises in the economy are assumed to be democratically organized cooperatives whose members hire management and workers, lease capital and land from the state, purchase commodities from and sell them to other cooperatives and individuals, and share profits and losses. This system is shown to be economically efficient given that the central planning board carries out its postulated four basic functions.

MICHAEL J. P. MAGILL, Ph.D. Brown 1970. On a general economic theory of motion.

A general economic theory of motion is presented, similar to the theory of motion in analytical mechanics. The *pure allocation equations of motion* are derived from a central principle: Fisher's *valuation principle* together with the *maximum principle* lead to the *dual equations of motion*. The present value function satisfying the Hamilton-Jacobi equation, unifies the theory. The possibility of *wave motion* is examined by associating a wave function with the economic system, the wave function satisfying a wave equation similar to Schrödinger's wave equation.

MUKUL K. MAJUNDAR, Ph.D. California (Berkeley) 1970. A general theory of efficiency prices for economies with infinitely many commodities.

JAMES MAK, Ph.D. Purdue 1970. Production, consumption, and the distribution of agricultural surpluses and deficits in Ohio: 1840–60.

The primary objective of this study was to formulate a model (algorithm) which could do two things. First, it could generate information which would tell us something about agricultural development in the antebellum American West—in particular, the agricul-

tural development of the old Northwest Territory. Second, the model could be used to generate data on the geographic distribution of food surpluses or deficits within the region. It was hoped that the achievement of this objective would induce others to use this model to study various aspects of regional development in the antebellum American West.

G. F. MATHEWSON, Ph.D. Stanford 1970. Economic patterns of information seeking.

A dynamic consumer decision model is constructed which emphasizes the role of accumulated information acquired at a cost through the media. Empirical tests are conducted on the time profile of resource allocation by the consumer to these informational ends. Consideration is given to the role of advertising in this informational activity and the effect of financial assets on the informational stocks held by the consumer.

HELMUT A. MERKLEIN, Ph.D. Texas A&M 1970. An evaluation of German contributions to the theory of distribution.

The object of this dissertation is to present a critique of recent distribution theories that have been published in the German language. The work of the following five authors is covered: G. Bombach, A. Stobbe, E. Preiser, W. Krelle, and E. Schneider. Since most of the material in this dissertation has never been presented in English, the principal models of these authors are first summarized in English. These summaries are then followed by a neoclassical critique.

ROBERT T. MICHAEL, Ph.D. Columbia 1969. The effect of education on efficiency in consumption.

The concept of human capital and its effect on market earnings has been studied extensively in the past decade, but little attention has been paid to its possible effects in nonmarket activities. In context of household production functions, this dissertation explores how human capital might affect the productivity of these functions, alter real income and relative prices, and affect consumer behavior. Empirically it investigates the effect of formal schooling on consumer expenditure patterns.

MARY J. OATES, Ph.D. Yale 1969. The role of the cotton textile industry in the economic development of the American Southeast: 1900–40.

NORMAN P. OBST, Ph.D. Purdue 1970. Stability of a competitive market with speculative behavior.

This paper investigates the stability of a two-commodity market in which trading occurs during scheduled, separate intervals of time. Using both differential and difference equations, the effect upon stability of changing the length of trading each day is examined. A new type of resource allocation process for organized markets is introduced as a foundation for discussion. The effects of speculation upon stability are also studied.

TERRENCE B. O'KEEFE, Ph.D. Purdue 1970. Variance dependent production decision rules.

ARNOLD H. PACKER, Ph.D. North Carolina (Chapel Hill) 1969. Adaptive modeling procedures.

R. EDWARD PARK, Ph.D. Princeton 1970. Transportation and efficient growth.

This paper discusses efficient growth in a series of models that explicitly incorporate spatial separation and costly transportation. All of the models are based on linear technologies of von Neumann type. Aspects of efficient behavior are explored in each of the models. The central results of the paper concern the way in which efficient frontiers evolve over time and the nature of efficient growth paths.

JACOB PAROUSH, Ph.D. Columbia 1969. Consumption technology and order relations in consumption.

A model of consumer behavior was constructed in the framework of which order relations in consumption can be explained as a result of efficient purchasing behavior. The model provides the possibility of making inferences about the structure of the consumption technology from empirical observations. The relationship between income and the consumer's stage on the scale of acquisition was established along with the technological complementarity relation among goods which perform an order of acquisition.

HARRY H. POSTNER, Ph.D. Minnesota 1970. Estimation of the elasticity of capital-labor substitution in postwar Canadian manufacturing industries.

Estimation proceeds in the context of a fifteen equation general model including a gross output production relationship, and demand equations for labor and capital input derived from an inequality constrained conditional expectation maximization decision hypothesis. The key requirements for manipulations within the model are the application of a Kuhn-Tucker theorem, and the assumption of separable partially reduced form versions of labor and capital supply equations. The estimation method is a generalized instrumental variable technique.

BRUCE M. PRITCHETT, Ph.D. Purdue 1970. A study of capital mobilization, the life insurance industry of the nineteenth century.

A financial history of all legal reserve companies of the nineteenth century which provides a case study of early intermediation. These companies effectively gathered funds but evidenced distortions in their investment practices. Southern and western regions consistently experienced net outflows of cash, and growth related industries received little direct investment during critical periods in their development, largely because of provincial legislation and management practices which returned no discernable benefit through decreased investment portfolio risks.

MARC J. ROBERTS, Ph.D. Harvard 1970. Models and theory in economics: an examination of the logical

status of economic theory with an application to welfare economics.

MARSHALL ROSE, Ph.D. Tulane 1969. Development and application of a time-constrained inventory model for reparable assets.

This thesis develops an inventory model for both reparable spares and end products, and incorporates the model into a theory for a particular class of firms. The firms considered operate a fleet of end products and service them periodically. The relationships between the quantity of spares and other repair resources, and the end-product service time are calculated. The appropriate conditions of optimality are derived for these decision variables and for the total number of end products to acquire. We obtain a specified operational level of end products at least cost.

PAUL H. RUBIN, Ph.D. Purdue 1970. A theoretical model of the diversification decision of the firm.

The firm is viewed as a collection of particular resources, i.e., resources worth more to the firm than their market value. Such resources can be used either for producing output or for training new resources. This assumption is incorporated in a programming model of the firm, which is then used to derive prediction statements about the rate and direction of diversification.

LAWRENCE D. SCHALL, Ph.D. Chicago 1969. Technological externalities and resource allocation.

This thesis is an examination of the conditions necessary for technological externalities and of the consequences of such externalities for resource allocation. A competitive equilibrium with technological externalities present in a given industry is compared with a Pareto optimum in terms of input use and final outputs of the economy. Resource allocation with technological externalities and imperfect competition is also considered. Deficiencies of accepted theory are corrected and new results are derived.

GEORGE E. SCHNEIDER, Ph.D. Notre Dame 1969. Adam Smith's theory of economic policy.

HOWARD SCHUMACHER, Ph.D. Catholic 1969. The decline of the Dutch coal mining industry and its effect on the economy of South Limberg.

ALLEN SINAI, Ph.D. Northwestern 1969. A theoretical and empirical study of fixed investment in the U.S. steel industry, 1947–65.

A general model of capital expenditures containing as special cases, capacity-accelerator, marginal efficiency of investment, and liquidity hypotheses is postulated and tested using annual data from 1947–65 for 45 steel firms in the United States. Better results are obtained, regardless of the structuring of the data, for the general hypothesis. It is concluded that future research in investment should focus on refinement of a general model rather than separate investigation of competing special cases.

HENRY L. STETTLER III, Ph.D. Purdue 1970. Growth and fluctuations in the antebellum textile industry: 1790–1860.

NEIL M. SWAN, Ph.D. Pennsylvania 1969. Inflation and the distribution of income.

The effects of inflation on the postwar distribution of income in the United States are examined, treating transfer incomes and market incomes separately. Most transfer incomes, notably OASDI benefits, are found to be determined in such a way that protection against inflation is almost complete. For the distribution of market incomes, tests indicated that inflation has no significant effect. The conclusion is drawn that the income redistributional effects of inflation are very small.

FEREYDOON TAFAZZOLI, Ph.D. Southern California 1970. Oskar Lange's contribution to the economic theory of socialism, with applications to Yugoslavia and Czechoslovakia.

After World War I, Ludwig von Mises, Friedrich von Hayek, and Lionel Robbins argued that socialism was incapable of attaining a rational economic calculation. In the 1930's Oskar Lange categorically rejected this contention. The purpose of this dissertation is to critically reexamine Lange's contributions to the economic theory of socialism and to assess the relevance of the Lange model(s) to the evolution and development of the Yugoslav and Czechoslovak economic systems.

CHARLES TAPIERO, Ph.D. New York 1970. Graphs: theoretic approach to powers.

DAVID G. TARR, Ph.D. Brown 1970. The stability theory of dynamic oligopoly under uncertainty.

This dissertation examines the conditions under which an oligopolistic market will converge to an equilibrium arbitrarily closely, when in any time period each oligopolist is uncertain about the amount of output its competitors will produce in the following time period. Each firm conceives of its competitors' output as a random variable having a probability distribution. Models are considered under general market conditions, and under the assumption of a linear market demand curve and quadratic cost curves for each firm. Under the latter assumptions the market is more likely to be stable: the fewer the number of firms; the faster marginal costs increase; the more elastic the demand curve; and the less the mean of the probability distribution changes when any firm's output level changes.

JACQUES E. TERNOY, Ph.D. Iowa (Ames) 1969. Cooperation and economic efficiency.

This study deals with a theoretical economic analysis of cooperative organizations. The analysis is restricted to considerations bearing directly and indirectly on efficiency at the levels of the firm and of a group of firms. The main objective is to investigate the economic reasons which theoretically explain the rapid

growth of cooperative organizations as substitutes for competitive organizations, as well as the inability of the cooperative system to supplant completely the competitive system in most economic areas.

PAUL W. THOMPSON, Ph.D. Duke 1970. An exposition and analysis of Marx's prediction of communism.

GEORGE W. TRIVOLI, Ph.D. Virginia 1970. A critical examination of J. K. Galbraith's industrial system hypothesis.

THEODORE TSUKAHARA, JR., Ph.D. Claremont School 1970. The behavioral foundations of the theory of consumer choice under uncertainty.

It is integral to the process of decision making under uncertainty for the individual to seek out information sources, gather data, interpret them, and decide either to continue the search or to make a choice. We have formalized this behavioral process as the transformation of information into useful form, i.e., the production of knowledge. Two models are developed, a continuous version of the production function and a mathematical programming version.

NANCY H. WILSON, Ph.D. California (Berkeley) 1970. Optimal planning of economic development in an open economy: a dynamic linear logarithmic model.

RICHARD D. WOLFF, Ph.D. Yale 1969. Economic aspects of British colonialism in Kenya: 1859–1930.

BENJAMIN WOLKOWITZ, Ph.D. Brown 1970. Homothetic production functions: a theoretical and empirical analysis.

By employing the relationship between observable economic variables (e.g., *MRS, K/L*, etc.), a class of production functions are specified which may have constant or variable elasticities of substitution. Homotheticity is incorporated in these classes so that an explicit representation of a production process where the returns to scale are functionally related to the output level is possible. These new production functions are empirically tested with U.S. manufacturing data (both cross-section and time-series).

DOUGLAS W. WOODS, Ph.D. Massachusetts Institute of Technology 1969. The price game.

GAVIN WRIGHT, Ph.D. Yale 1969. The economics of cotton in the antebellum South.

KENNETH C. YOUNG, Ph.D. Purdue 1970. Stochastic and non-stochastic models of labor supply.

The thesis generalizes the traditional theory of labor supply to include the effect of risk. It includes models which permit the analysis of salary payments vs. piece-work and the effects of risk on the allocation of labor in on-the-top choices such as those faced by salesmen, entrepreneurs, and executives. An appendix shows how the computer can be used in the analytic solution of problems pertaining to optimization models.

JEFFREY F. ZABLER, Ph.D. Pennsylvania 1970. A microeconomic study of iron manufacture, 1800–30.

This dissertation seeks to clarify the course of American antebellum economic development by building an operational model of the pig iron industry in the 1800–30 period and analyzing resulting patterns in light of current views of U.S. growth ranging from Walt Rostow's "take-off" to Paul David's gradualism. From manuscript sources a highly detailed picture of actual costs, capital, output, size, technology, profit levels, and a variety of institutional and market behavior patterns is worked into a common theme.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

CAROL M. ANDERSON, Ph.D. Boston College 1970. Contributions of agriculture to Nigerian economic development.

This study analyzes the contributions of the agricultural sector to Nigerian economic development between 1950 and 1964. Nigeria is shown to have been relatively self-sufficient in food production. Lower prices to domestic producers than were received for agricultural exports helped to restrict consumption. Foreign exchange earnings accumulated mostly in the earlier years were used to finance the imports needed for economic development.

ALDO A. ARNAUDO, Ph.D. Yale 1970. The determination of the rate of change of manufacturing prices under inflationary conditions.

AZMAN AZIS, Ph.D. Pennsylvania State 1970. Projections of Mexico's future requirements of base and precious metals and estimates of domestic potential supply.

HOWARD N. BARNUM, Ph.D. California (Berkeley) 1969. A model of the market for foodgrains in India 1948–64.

RONALD F. BEARDEN, Ph.D. Houston 1970. The causal nature of financial development.

Several theorists have suggested economic progress may be stimulated by financial means. Traditional thinking generally describes the financial function as passive. This study empirically tests hypotheses concerning the interrelationship of real economic growth and financial development during the 1940–60 accelerated growth period in Mexico. The results indicate a financial system can be supply-leading, that is, current economic growth relates more significantly to financial development in the preceding growth period. There was a rise-peak-fall pattern to this relationship partic-

ularly evident in the development of public nonmonetary financial institutions.

Ivan Bello, Ph.D. Southern California 1969. A simple macroeconomic model for the Chilean economy.

This dissertation derives a simple macroeconomic model for the Chilean economy. The model consists of a consumption function, a production function, and three definitional equations. National data for period 1940-65 were utilized in estimating the relationships. Excellent fits were obtained for all relationships. The simplicity of the model was determined by the availability of data. Better data would, undoubtedly, improve the predictive capacity of the model.

Michael H. Best, Ph.D. Oregon 1969. Determinants of tax performance in developing countries: the case of Guatemala.

Eleanor Birch, Ph.D. Iowa 1969. A permanent demand analysis of investment.

Tyrone Black, Ph.D. Tulane 1970. The dominant causes of long-term shifts in the industrial composition of output in the United States: an analysis and classification.

A model is developed that identifies which side of the market, the supply-side or the demand-side, was the governing influence on output change for any output category over a given time period. The model is used to identify the governing influence on output change for a number of commodity classifications of the United States economy over the period 1899 to 1965. The results are employed to explain the causes of structural change predicted by the Fisher-Clark thesis.

Ernest Bloch, Ph.D. New York 1970. An economic analysis of recent grain trends in the developing countries.

This study seeks to aid in answering the question of whether a downturn in per capita food production in the developing countries began in the 1960's. For this purpose grain trends from 1957–59 to 1963–65 are analyzed. The impact of the introduction of high-yielding wheat and rice varieties in Asia in the mid-1960's and their role in future output are also considered.

Barry P. Bosworth, Ph.D. Michigan 1969. Alternative models of investment behavior.

Leonard G. Bower, Ph.D. Duke 1969. Population growth, economic growth, and family planning programs in less developed countries.

Colin I. Bradford, Jr., Ph.D. Columbia 1970. Planning, public investment allocations, and intersectoral relations in the Chilean economy.

A hypothesis explaining the low degree of implementation of the public investment planning strategy in Chile from 1961 to 1967 is that public investment was used in practice to achieve short-run economic goals. The effects of alternative investment strategies on three short-term economic goal conditions are compared. A policy model utilizing input-output analysis is presented which relates the sectoral allocation of investment to appropriate endogenous variables. Interest group pressures are also analyzed.

Jay B. Bright, Ph.D. Houston 1969. A study of certain aspects of the productivity of industrial labor in independent India.

The dissertation tests a hypothesis which states that an inexperienced industrial labor force, without a prolonged industrial learning process behind it, prevented the launching of an industrial revolution in India during the decade of the 1950's. Productivity of industrial labor in India and Britain for different sets and combinations of established and recent industries was studied. Results of multiple and rank correlations lent only partial support to the hypothesis.

Thomas F. Cargill, Ph.D. California (Davis). A spectral analysis of the wage-lag hypothesis.

James Carter, Ph.D. Oregon 1969. The net cost of Soviet foreign aid: 1955–65.

Richard D. Carter, Ph.D. Ohio State 1969. The effect of a capital development project on the national income of a developing country; Malaysia.

Robert Cherry, Ph.D. Kansas 1969. Growth in a two-country two-commodity competitive world.

Pat Choate, Ph.D. Oklahoma 1969. An economic development program for Oklahoma.

This study describes the magnitude of the under-development in Oklahoma, alternative solutions, a strategy to eliminate this under-development, specific research, specific public actions and specific public expenditures to eliminate this regional economic stagnation. The proposed strategy, which is now being implemented, focuses on creating new industrial jobs by using public investments as a catalyst for private risk capital. The strategy and the proposed projects are detailed to the extent that they are ready for implementation.

Hakchung J. Choo, Ph.D. Clark 1970. An interindustrial analysis of development strategy and performance: a case study of the Korean first five-year plan, 1962–66.

Development strategy and performance of the Korean first five-year plan, 1962–66, are examined empirically and by means of interindustrial analysis. Feasibility of the plan and consistency in resource allocation with respect to development objectives are tested. The impacts of development efforts on the Korean industrial structure and technical transformation are also examined utilizing the 1960, 1963, and 1966 input-output tables. Lewis' three criteria are applied for an overall appraisal of planning and performance of the plan.

PETER K. CLARK, Ph.D. Harvard 1970. A subordinated stochastic process model of cotton futures prices.

This thesis proposes a subordinated stochastic process model for the price-series of cotton futures. The subordinated process $X(T(t))$ is fitted to price index for cotton futures. Trading volume is taken as an instrument for the increments of the operational time process $T(t)$, and it is shown that a subordinated process with delta $T(t)$ distributed lognormally and delta $X(t)$ distributed normally fits the data better than stable model. The serial dependence of volume is shown to explain the X-shaped transition matrix of price changes.

ANTHONY T. CLUFF, Ph.D. George Washington 1970. Prices, unit labor costs, and profits: an examination of Wesley C. Mitchell's business cycle theory for the period, 1947–69.

The relationship between prices, unit labor costs, and profits outlined in Wesley C. Mitchell's business cycle theory was examined for the postwar period using data for the private nonfarm economy. The cyclical behavior of these variables generally conformed to Mitchell's theory. Profit margins declined when unit labor costs increased rapidly, prior to a downturn in economic activity. A relationship was also shown to exist between capacity utilization rates and changes in labor productivity.

ELIZABETH F. CROOK, Ph.D. Fletcher School 1970. Political development as an objective of United States foreign assistance: Title IX of the 1966 foreign assistance act.

BLAIR C. CURRIE, Ph.D. Rochester 1969. Agricultural output and consumption in mainland China, 1929–33 and 1957.

This study appraises the validity of official output data for food grains, soybeans, and cotton in mainland China for 1957. Output totals for those crops are reconstructed from disaggregative data. These 1957 estimates agree quite closely with the official output figures. A comparison is then made, by province, of prewar output and consumption estimates with similar estimates for 1957. Both comparisons imply that official 1957 output totals are understatements of actual output.

GUNVANTRAI M. DESAI, Ph.D. Cornell 1969. Growth of fertilizer use in Indian agriculture: past trends and future demand.

The study analyzes past fertilizer use in order to estimate future demand. Rates of fertilizer application have been significantly below standard recommendations. The proportion of area fertilized has increased substantially while rates of application changed little. As nitrogen use approaches 2 million tons per year, rapid growth in demand will require improvement in crop varieties and irrigation. The first continues technological change on presently irrigated areas, the second broadens the base of use.

MICHAEL R. DOHAN, Ph.D. Massachusetts Institute of Technology 1969. Soviet foreign trade in the NEP economy and Soviet industrialization strategy.

The study examines Soviet foreign trade, foreign trade plans, and Soviet economic policy during the period 1921–28. It focuses on the causes of export stagnation and the effects of the import constraint on economic recovery and on the decisions leading to industrialization. New statistics for Soviet foreign trade for 1913–1940 include foreign trade plans, price and volume indexes, foreign reserves and foreign debt, import-consumption ratios, and export-output ratios.

EL-SAYED A. DOHIA, Ph.D. Colorado 1969. The external debt and debt-servicing problems of the developing countries with special reference to the UAR.

EDWARD DOLAN, Ph.D. Yale 1969. The evolution of the Soviet planning system.

STANLEY W. DRISKELL, Ph.D. California (Berkeley) 1970. An analysis of economic causes and results of labor migration in Tanganyika.

ROBERTO P. ECHEVERRIA, Ph.D. Cornell 1969. The effect of agricultural price policies on intersectoral income transfers.

This study aims to isolate the effects of variations in relative prices that bring about changes in the proportion of national income obtained by different economic groups in a country. Chile is used as a case study. The groups for agriculture are: large producers, small producers, tenant laborers and agricultural workers; for other sectors: non-agricultural producers, non-agricultural workers and the rest of the world. The analysis identified the magnitude of benefits to some economic groups and the costs to others of changes in relative prices between productive sectors.

FAIKA M. M. EL-RIFAIE, Ph.D. Colorado 1969. Direction of trade and economic development in the UAR: 1950–68.

ALEJANDRO FOXLEY, Ph.D. Wisconsin (Madison) 1970. Structural disequilibrium and alternative growth patterns for the Chilean economy, 1970–80.

A multisectoral programming model is used to describe alternative growth and employment patterns for the Chilean economy in two terminal years (1975 and 1980). Particular emphasis is given to the analysis of import-substitution and export-promotion strategies and their effect on growth and employment; to the study of substitution between internal and external savings; and to the social cost implied by different sectoral employment policies. Results from the multisectoral model are fed into an aggregate macro-model which is used to provide some degree of intertemporal consistency to the terminal year projections.

KWOK-KWAN FUNG, Ph.D. Harvard 1970. Growth and utilization of the labor force in Taiwan, 1956–66.

This is a case study of labor utilization in less developed countries undergoing rapid population and economic growth. The labor force was absorbed with rapid relative shifts towards non-agricultural and services employment propelled by the manufacturing sector, while the extension of formal education and the growth of non-family employment lead to a large decline in male participation and a large increase in female participation in the 15–19 age group.

M. A. FUTAYYEH, Ph.D. Oklahoma 1970. Sales tax for economic development: Syria as a case study.

This study examines in detail the internal methods of taxation for development. It is shown that the present financial resources are not adequate to finance the development programs and concurrently meet the administrative expenses. The study recommends the substitution of a sales tax for an income tax to secure the necessary revenue to finance the development programs and insure a steady economic growth.

FAY E. GRECKEL, Ph.D. Indiana 1969. Import substitution in underdeveloped nations and regions: the Mexican experience.

This dissertation presents a critical analysis of the theoretical basis for import substitution and an empirical study of import substitution in Mexico between 1940 and 1960. The empirical study of Mexico illustrates various implications of both the national and regional analysis. An analysis is included of the linkage effects associated with import substituting industrialization and an examination of the effects of import substitution on regional development.

ANGELICO A. GROPPELLI, Ph.D. New York 1970. The growth process in the computer industry.

This thesis analyzes the role played by technology, competition, and demand in promoting growth of computers. Product cycles are developed and indicate how demand changed over three generations of computers. Technological change, which contributed to lower costs of product improvement and climbing profits, is largely responsible for growth in the industry. This study focuses attention on the interaction of technological change and economics in explaining the growth process of an industry.

STEPHEN E. GUISINGER, Ph.D. Harvard 1970. Effective protection, resource allocation, and the characteristics of protected industries: a case study of Pakistan, 1963–64.

This thesis studies effective protection from both a theoretical and empirical perspective with the purpose of determining the nature of the relationships between effective rates and key industrial characteristics, most notably indicators of interindustry factor movements. The chapters deal with the history of the concept of effective protection; the development of a simple geometric model of effective protection; a review of three studies of effective protection in Pakistan; sensitivity analyses on the structure of protective rates; statistical tests between effective rates, and other key industrial characteristics; and a summary with policy recommendations.

UMESH C. GULATI, Ph.D. Virginia 1969. Some aspects of economic policy in India in recent years and their effects on the prospects for private enterprise.

EDWARD H. HANIS, Ph.D. Indiana 1969. A theoretical investigation of embodied and disembodied innovation, economic growth, and structural unemployment.

This investigation finds that technical progress can be usefully studied within the framework of a growth model in which both disembodied and capital-embodied innovations occur. Using modified vintage-type Cobb-Douglas production function forms to represent both types of progress, an equilibrium aggregate output path is obtained wherein the growth rate depends not only upon the respective rates of embodied and disembodied innovation, but also upon the mix of these two types of progress. This growth model is developed into a theory of structural unemployment based on worker obsolescence. The model is used to develop rules for public expenditure on research and development and retraining.

CLELL G. HARRAL, Ph.D. Rochester 1969. The social costs of highway transport and the coordination of transport development in Eastern India.

Highway freight transport costs, including highway construction and maintenance and vehicle operating costs, have been estimated in a social cost framework suitable for comparative analyses among modes. Long-run average costs are found to be virtually constant beyond the extreme lowest traffic volumes. A comparison with railway costs and services suggests that greater highway competition would be socially desirable. The superior highway technology in handling LCL shipments, speed, flexibility and dependability of scheduling are important.

JOAN G. HAWORTH, Ph.D. Oregon 1970. Distributed lag analysis of the impact of monetary policy on the lumber industry of the Pacific Northwest.

ROLF G. H. HENRIKSSON, Ph.D. Northwestern 1969. An interpretation of the significance of emigration for the growth of economic welfare in Sweden, 1860–1910.

By means of population projection and a wage-model, a case is established for considering the emigration a factor contributing to the growth of economic welfare in Sweden.

REED HERTFORD, Ph.D. Chicago 1970. Sources of change in Mexican agricultural production, 1940–65.

The purpose is to explain the 1940–65 expansion in farm output, and to link irrigation and land reform developments to changes in the employment and returns of particular inputs. Production functions are identified, using 3,000 observations from county-level

summaries of the 1960 Mexican Agricultural Census. Separate functions are estimated for counties affected by land reform and irrigation. Parameters from these relations are used to calculate input contributions to output through time.

PETER J. HILL, Ph.D. Chicago 1970. The economic impact of immigration into the United States.

This study is concerned with the large influx of people into the United States from 1840 through 1920 and with their impact on the economic growth of the country. Research indicates that the foreign born were quite similar to the native born in earnings and occupational position and that the immigrants' effect on the economy has often been overrated.

CHARLES HOLLOWAY, Ph.D. California (Los Angeles) 1969. A mathematical programming approach to decision processes for complex operational systems with the aggregate planning problem as an example.

TIEN-TUNG HSUEH, Ph.D. Colorado 1969. An econometric model for Taiwan economic development.

SHENG CHENG HU, Ph.D. Rochester 1970. Technical progress and optimal growth.

This dissertation considers the optimal programs of capital accummulation and research in a neoclassical growth model with endogeneous technical progress. It shows that the optimal path depends not only on the initial capital-labor ratio, but also on the initial level of capital efficiency. In the long run, technical progress exhibits Harrod neutrality; the equilibrium growth rate depends on the discount rate, and the invention possibility frontier, but not on the production function.

ABDUL S. IDRISI, Ph.D. Syracuse 1970. Economic development with unlimited supplies of capital.

JOHN W. ISBISTER, Ph.D. Princeton 1969. The growth of employment in Mexico.

The dissertation investigates whether sufficient decent productive jobs are likely to be provided for Mexico's rapidly growing labor force in the future. It concludes that employment in Mexico has grown comparatively fast in recent years; a principal cause has been the slow growth of real wages. Nevertheless the labor force is growing so quickly that it is not certain that supply and demand will converge in the future unless specific policies are undertaken to ensure that they do.

HENRY E. JAKUBIAK, Ph.D. Harvard 1970. Model of structural inflation in a developing economy.

LOVELL R. JARVIS, Ph.D. Massachusetts Institute of Technology 1969. Supply response in the cattle industry: the Argentine case, 1937–38 and 1966–67.

The thesis develops and estimates an econometric model for the Argentine cattle sector. Cattle are considered to be capital goods which are held by producers as long as their capital value exceeds their slaughter value. Capital theoretic micro-models are developed to provide a basis for the estimated macro-model. Disaggregated herd stocks are constructed. The significance and implications of the results for the Argentine economy are discussed.

ALBERT E. JOHNSON, Ph.D. New York 1970. The growth and prospects of the foreign government supported wholly or predominantly tertiary economy with specific reference to Gibraltar.

J. FREMON JONES, JR., Ph.D. Florida 1970. Tourism as a tool for economic development with specific reference to the countries of Jamaica, Trinidad, and Guyana.

MOHAMMAD S. KALLA, Ph.D. American 1969. The role of foreign trade in the economic development of Syria, 1831–1914.

The role of trade in the development of Syria (and the term is used here in the broad, historical sense) is assessed in the light of relevant theoretical propositions. The British commercial and consular reports form the main source of the statistical basis for this work. Trade and population statistics are organized in two appendixes.

NORMAN M. KAPLAN, Ph.D. Chicago 1970. Studies in Soviet capital formation and economic growth.

The thesis examines several related questions: some methodological and informational problems of Soviet data on capital formation; the criteria for investment choice in Soviet theory; the extent to which U.S.-USSR differences in growth can be accounted for by differences in the rate and pattern of capital formation; the extent to which the recent deceleration in Soviet output growth can be accounted for by changes in the growth of inputs.

MOHAMMAD I. KHAN, Ph.D. Stanford 1969. Economics of food consumption in developing countries: a case study of Pakistan.

This study is centered around comparative analysis of demand and supply for food in Pakistan. The projection of future demand on alternative assumptions regarding population and income growth indicates great responsibility on Pakistan's agriculture to produce food for domestic consumption. But the recent innovations introduced in West Pakistan's agriculture have brought considerable expansion in food grain supply. Pakistan must now seriously think of the direction of agriculture in the stage beyond self-sufficiency in staple food.

DAVID J. KLOCK, Ph.D. Columbia 1969. The impact of direct trade controls and an overvalued exchange rate on factor proportions in manufacturing: the case of the Philippines, 1949–62.

MARTIN J. KOHN, Ph.D. Yale 1970. The stock of unfinished construction in the USSR, 1950–65; an efficiency problem in a centrally planned economy.

OREST KOROPECKY, Ph.D. Columbia 1969. The uniform development of manufacturing and the character of production.

The dissertation is primarily concerned with the fact that the development of manufacturers is uniform rather than various. I conclude that changes in the structure of manufacturing are in the direction of industries with the highest skill proportions becoming those with the largest shares of value-added. They are not simultaneously in the direction of industries with the highest capital/labor intensities becoming those with the largest shares of value-added. But, a process that overall is not increasingly capital intensive nevertheless tends to produce a larger and larger relative share of capital goods.

JAN A. KREGEL, Ph.D. Rutgers 1970. The role of the rate of profits and distribution theory in models of long-run economic growth.

The necessary requirements for a model of economic growth are found to be a theory of profit rates and a theory of distribution. It is also shown that the profit rate must be determinate if any distributional statements are to be made. Existing models of neoclassical and Keynesian origin are analyzed subject to these criteria. The Keynesian model is then extended to include financial aspects of distribution and alternative savings assumptions.

PATRICIA KUWAYAMA, Ph.D. New York 1970. Effective tariff protection in Japan.

LUIS LANDAU, Ph.D. Harvard 1970. Differences in saving ratios among Latin American countries.

Several models of economic development imply that the aggregate saving ratio is a non-linear increasing function of the level of per capita income among poorer countries. In this study, statistical tests for Latin American countries uphold the non-linearity hypothesis. The saving functions obtained from cross-sections are found to be equal to those resulting from pooling of time-series. Separate tests show the price-elasticity of investment to be less than unity.

DAE SUNG LEE, Ph.D. Massachusetts 1970. International trade in the economic development of the Korean economy.

EDWIN R. LIM, Ph.D. Harvard 1970. An export-propelled model of growth.

TSUNG-HUA LIU, Ph.D. Toronto 1969. An economic growth model for a dualistic underdeveloped economy.

This thesis is an investigation of theoretical aspects of underdeveloped economies within the framework of a dualistic model. After a critical review of existing models, an alternative model is introduced. The sufficient conditions for the existence and uniqueness of the short-run equilibrium solutions are given; the stability of the long-run equilibrium paths of the system are studied. Optimal policies which could lead the system out of the "vicious circle" are also discussed.

WILLIAM F. LOTT, Ph.D. North Carolina State, 1969. The effect of demographic characteristics and the interest rate on the consumption function of an economy over time.

The objectives were to test whether demographic characteristics (structural change) can be ignored in estimating the aggregate average propensity to consume, and to obtain an estimate of the interest rate effect in the context of a Friedman-type consumption function.

GARY A. LYNCH, Ph.D. Washington State 1970. *Ex ante* and *ex post* approaches to inventory fluctuations: an investigation of business behavior under uncertainty.

WILLIAM W. MCCORMICK, Ph.D. Colorado 1970. A self-generating model of long swings for the U.S. economy: 1960–70.

JANET M. MCKEE, Ph.D. California (Berkeley) 1970. Industrial protection in the republic of Kenya.

CALLISTO E. MADAVO, Ph.D. Notre Dame 1969. Dualism in the Southern African economy.

RITA MALDONADO, Ph.D. New York 1970. Financial institutions and economic development in Puerto Rico: an historical and econometric analysis.

MICHAEL E. MANOVE, Ph.D. Massachusetts Institute of Technology 1969. A model of administrative economic planning and plan execution in Soviet-type economies.

Several models are developed which, taken together, form an abstract description of procedures that could be used by a Soviet-type administrative apparatus to plan and run an economy. One model is used to examine the possibility, in principle, of achieving consistency in short-term plans created by Soviet-type planning procedures. Another model provides an algorithm for determining optimal nonprice rationing procedures for intermediate goods. These rationing procedures could be used by an administrative apparatus to handle shortages that arise during the execution of a short-term plan.

PHILIP M. MATHEW, Ph.D. McGill 1969. The economics of rubber plantations in India.

This is an economic evaluation of the feasibility of government investment in the rubber plantation industry in India. The benefit-cost methodology was used. Cash flows were discounted at rates from 5 to 15 percent, with emphasis on a discount rate of 10 percent, which is approximately the marginal productivity of capital. The optimum project size is 5,000 acres, with planning horizons of 37 years, 32 years, and 27 years. Direct benefit-cost findings, including the internal rate of return (e.g., 14.94 percent for 37

years) and present worth, indicate that rubber plantations are a worthwhile public sector investment. This conclusion is further supported by an appraisal of important secondary benefits.

RICHARD L. MEYER, Ph.D. Cornell 1970. Debt repayment capacity of the Chilean agrarian reform beneficiaries.

The objectives of this study are to record the development of Chilean agrarian reform, to describe the present method of implementing reform, and to project the debt repayment capacity of beneficiaries with present and potential farm income. Assuming maintenance of present farm population and current levels of consumption by workers, half of the asentamientos studied generate sufficient income to pay projected debt installments. If the alternatives of improved management were adopted, labor use would drop to 70 or 80 percent of total man-days currently employed, gross value of farm production would increase by at least 25 percent, operating capital requirements would rise more than 13 percent, and debts could be serviced.

JOHN J. MINGO, Ph.D. Brown 1970. Labor mobility, capital importation, and economic growth: the case of postwar Puerto Rico.

LEONARD J. MIRMAN, Ph.D. Rochester 1970. Optimal growth and uncertainty.

RUSSELL M. MOORE, Ph.D. Fletcher School 1970. The role of extrazonally controlled multinational corporations in the process of establishing a regional Latin American automotive industry: a case study of Brazil.

BARBARA L. MOORES, Ph.D. Michigan 1969. Simulation and policy analysis with an economic model.

W. DOUGLAS MORGAN, Ph.D. California (Berkeley) 1969. The determinates of commercial investment behavior, 1949–64.

ISKANDAR M. NAJJAR, Ph.D. Indiana 1969. The development of a one resource economy: a case study of Kuwait.

A study of the feasibility of creating a viable and prosperous Kuwait economy from its present dependence on oil.

MORRIS R. NORMAN, Ph.D. Pennsylvania 1969. The Great Depression and what might have been: an econometric model simulation study.

A revised version of the Klein-Goldberger model is used as the framework for this study. The coefficients of the model are estimated by ordinary least squares, two-stage least squares based on principal components, and full information maximum likelihood. The potential effectiveness of modern anticyclical policies on the economy during the Great Depression are assessed. The stabilizing influence of modern tax structure and income transfer mechanisms are examined.

CHARLES NTAMERE, Ph.D. Notre Dame 1970. The feasibility of an iron and steel industry in Nigeria (and Biafra).

FRED OBERMILLER, Ph.D. Missouri 1969. Factors associated with agricultural development and growth in Latin America.

OSCAR OZAETA, Ph.D. Indiana 1969. An econometric study of the private consumption sector in a developing economy: the case of Spain.

MANOUCHER PARVIN, Ph.D. Columbia 1969. Technological adaptation and the rate of per capita income growth: an econometric approach.

Invention-innovation leading to increased productivity is assumed to be on the average a more costly activity than the adaptation of technology if an identical result is to be obtained. The rate of technological adaptation is determined by the technological gap, the stock of human capital per capita, and the rate of accumulation of capital. These factors are incorporated in a growth model whose further properties are derived and studied. The results of the cross-sectional multifactor regression analysis explaining the income growth differential of 46 countries, strongly support the former theory.

GEORGE F. PATRICK, Ph.D. Purdue 1970. Education and agricultural development in eastern Brazil.

The firm-level production function, with formal schooling and extension activities included as explicit variables, was used to estimate returns to these educational activities. Surveys of farmers in five areas representing different levels of agricultural modernization furnished information necessary to estimate returns to educational activities, costs of participating in extension, students' earnings foregone, and direct costs of formal schooling. Information on public costs of education activities was derived from secondary sources.

BRIAN R. PAYNE, Ph.D. California (Berkeley) 1969. An economic analysis of alternative rates of investment in national forest road construction: the North Umpqua case.

Determined that the economically efficient rate of timber-access road construction for a specific national forest unit is at least as slow as the present rate, given the allowable annual timber harvest. Suggested that faster roadbuilding on the national forests is undesirable unless the allowable cut restraint is relaxed or unless substantially greater timber or nontimber benefits are generated than were foreseen for the North Umpqua unit.

JOHN D. PILGRIM, Ph.D. Vanderbilt 1969. The upper turning point of 1920.

FRANK J. POPER, Ph.D. New York 1970. A critical evaluation of the empirical evidence underlying the relationship between hours of work and labor productivity.

Recent studies on the sources of economic growth have sometimes asserted that a positive and diminishing productivity offset is a normal consequence of secularly declining working hours over a certain range in the length of the average work week. The present study critically evaluates existing empirical evidence on this relation, or more generally, on the relation between changes in the length of the work week and their impact on labor productivity.

MOHAMMED A. RABIE, Ph.D. Houston 1970. The impact of the Aswan high dam on the economic development of the United Arab Republic.

The dam's contribution to agriculture was examined in terms of the expansion in the cultivated area, in the crop area, and the expected increase in the income of the agricultural sector. Its contribution to industry was examined in terms of the readily available hydroelectric power and in terms of the potential expansion in Egypt's industrial sector. In an effort to evaluate its overall impact on Egypt's economy and society, some other related factors have been considered: its effect on Egypt's national income, employment, navigation, manpower training, and potential social change.

MIGUEL A. RAMIREZ-PEREZ, Ph.D. Rutgers 1970. Functional income distribution in Puerto Rico: 1947–66.

This dissertation reexamines the definitions, measurement and evaluation of factor income shares in Puerto Rico in the light of the conventional net national income approach, versus a revised gross domestic income approach to functional income distribution. The superiority of the revised approach is firmly established in terms of its more realistic conceptual and analytical methodology. It yields drastically different distributive results in the analysis of Puerto Rico's predominantly open, rapidly growing industrial economy.

CHARLES T. RATCLIFFE, Ph.D. California (Berkeley) 1969. Tax policy and investment behavior in postwar Japan.

CARL A. RISKIN, Ph.D. California (Berkeley) 1969. Local industry in Chinese economic development 1950–57: the case of Kwang Tung province.

SHERMAN ROBINSON, Ph.D. Harvard 1970. Aggregate production functions and growth models in economic development: a cross-section study.

This dissertation is a cross-section study of 39 less-developed countries using regression analysis and theoretical models based on the aggregate production function and neoclassical growth models. A dynamic two-sector model is developed which includes the effect on growth of the rate of structural change and is also extended to include a foreign exchange constraint. Sources of growth and residual analyses are done and comparisons are made with time-series studies.

RITA RODRIGUEZ-MEDEROS, Ph.D. New York 1969. Fluctuations in United States imports 1948–66: economic analysis and econometric model.

MANUEL ROMAN, Ph.D. New School 1970. The limits of economic growth: the case of Spain 1959–67.

The thesis examines the development experience of Spain in the 1960's, focusing upon the system of goods and factor exchanges with Western Europe that followed devaluation of the peseta in 1959. The major constraints to economic development are then identified as the irrational organization of agriculture and the inflationary process which led to a decline in the rate of profits.

ODED RUDAWSKY, Ph.D. Pennsylvania State 1969. The contribution of a domestic cement industry to economic growth in developing countries.

YOUSSIF S. A. SABAH, Ph.D. Fletcher School 1970. Economic transformation of Kuwait: A study of a dualistic economy with capital surplus.

EISUKE SAKAKIBARA, Ph.D. Michigan 1969. Optimal growth and stabilization policy.

ALAN K. SEVERN, Ph.D. Pennsylvania 1969. Short-run investment and financial behavior of U.S. direct investors in manufacturing.

Short-run changes in foreign and domestic investment, dividends, and capital outflows less repatriated earnings are studied in a simultaneous-equation model. Firm intercepts explain more variation than do annual variations in sales and internal funds. Primary causation among the endogenous variables was from investment to dividends, and from these variables to balance-of-payments flows. Simulations showed little net effect of domestic conditions in the U.S. balance of payments via direct investment.

ERFAN A. SHAFEY, Ph.D. Southern California 1969. Development planning: a systems approach.

This study focuses on the processes of national economic development and development planning in less developed countries. Economic development is formulated in terms of interactions among societal subsystems. As a derivative of this formulation, development planning is conceived as the steering of societal subsystems and their interactions over time. The analysis which concentrates on the processes of optimization, innovation, and integration, emphasizes the complementary and competitive relationship between economic programming and modernizing ideology.

JOHN A. SHAW, Ph.D. Purdue 1969. The 1937 business cycle.

BONG JU SHIN, Ph.D. Ohio State 1969. Inflation and economic growth: An empirical study based on the Korean experience, 1948–69.

ROGER B. SKURSKI, Ph.D. Wisconsin (Madison) 1970. The distribution of consumer goods in the Soviet Union.

This study analyzes the domestic trade sector of the USSR as it has reacted to the development of the buyers' market and the rise of the Soviet consumer. The hypothesis that the traditional centrally planned, centrally administered system of consumer goods distribution in these conditions is less effective in terms of efficiency and consumer satisfaction than a more decentralized system is strongly supported by the evidence of this study. It was found that this problem has been recognized in the Soviet Union, and that adjustments have been taking place over the last several years on a very gradual basis.

KALBURGI SRINIVAS, Ph.D. California (Los Angeles) 1970. A computer simulation model of Newcomb's consistency theory: A case in theory development.

HERMAN STAROBIN, Ph.D. New York 1969. Czechoslovakia: 1948–68; a case study in socialist economic and political development.

The communist seizure of power in Czechoslovakia in 1948 is viewed in terms of the country as a supply base for capital goods for the Soviet Union and its political allies. After almost two decades of performing this function, crisis conditions developed in the Czechoslovak economy. Lack of sensitivity of the country's highly centralized system brought the economy to the brink of collapse. Soviet dissatisfaction with Czechoslovak communist attempts to deal with the causal factors of the crisis resulted in invasion and occupation.

LEONARD D. STEED, Ph.D. Columbia 1969. The nature of economic fluctuations in Argentina: an econometric study.

This dissertation attempts to determine the nature of the cyclical fluctuations of the Argentine economy and to evaluate the effectiveness of different stabilization policies. For this purpose an aggregate econometric model is fitted to Argentine data and a combined use of computer simulation techniques and spectral analysis is made to analyze both the stochastic properties of the fitted model and the relative effectiveness of different decision rules designed to stabilize the economy.

CHEN SUN, Ph.D. Oklahoma 1970. A multisectoral analysis of the business cycle.

This study is in an attempt to build a business cycle model on the bases of received microeconomic principles. It starts with a reconsideration of the acceleration principle in terms of profit maximization. Price-cost analysis is used throughout the work to explain the cumulative process: The downturn, and the upturn of the cycle. Finally, the validity of the model is tested by reference to historical data.

LEON P. SYDOR, Ph.D. Princeton 1970. National shares in St. Lawrence Seaway benefits.

Benefit-cost techniques are employed in an *ex post* evaluation of the St. Lawrence Seaway. Theoretical measures of national benefit shares are derived from partial-equilibrium models for wheat, corn, and iron ore. Primary benefits are then estimated and allocated among Canada, the United States, and countries in the rest of the world. After project costs are introduced, yield rates are calculated. The results suggest that the Seaway has been worthwhile.

MAHFOUZ E. TADROS, Ph.D. Cornell 1969. A game theoretic model for farm planning under uncertainty: An application to a cash crop farm in New York State.

The proposed model utilizes the theory of games and linear programming. A deterministic linear programming model for planning farms is employed to derive the strategies and the payoff matrix for the proposed game. Decision criteria that have been suggested for games against nature are applicable for that game. An empirical evaluation is presented by applying the game model to an 800 acre vegetable and other cash crop farm in New York State.

HERBERT M. THOMPSON, Jr., Ph.D. Colorado 1969. A study of economic transition.

CHARLES P. TIMMER, Ph.D. Harvard 1970. On measuring technical efficiency.

This thesis measures technical efficiency relative to an efficient production function and attempts to explain the inefficiencies in terms of managerial characteristics. The frontier production function for U.S. agriculture at the state level was both deterministically and probabilistically fit by single-signed least lines to a Cobb-Douglas function. The fits were excellent, but only small differences in technical efficiency were measured because six factors of production were included in the function.

YIANNIS VENIERIS, Ph.D. Oregon 1969. An investigation in the behavior of price and inventory movements in the postwar U.S. economy.

DANIEL P. VILLANUEVA, Ph.D. Wisconsin (Madison) 1970. Monetary and fiscal policies in aggregate growth models with endogenous technical change.

Variable amounts of the fully employed capital and labor are allocated to the production of new labor (in efficiency units); new capital and new labor are the joint growth mechanisms by which output per worker (per unit of population) may be increased. The short-run and long-run effects of monetary and fiscal policies on capital accumulation and growth are analyzed. The implications of the models for the growth rate of an economy are tested.

HOWARD M. WACHTEL, Ph.D. Michigan 1969. Workers' management and wage differentials in Yugoslavia.

ALAN R. WATERS, Ph.D. Rice 1969. The cost structure of the Kenya coffee industry.

The thesis examines the development and the present structure of the Kenya coffee industry, and makes estimates based upon two major sample surveys in Kenya of the opportunity cost of coffee production for both plantations and small holdings. It discusses alternative explanations for the cost findings, and evaluates a policy proposal for Kenya to maximize net returns to producing coffee under the overall constraints imposed by the International Coffee Agreement.

PETER J. WATRY, JR., Ph.D. Missouri (Columbia) 1970. The economic development of Baja California, Mexico.

The purpose of this dissertation was to examine the growth of exports from the state of Baja California since 1900 and to show how these exports have affected and stimulated economic development in the state. The three exports were cotton from the Mexicali Valley, American tourism in Tijuana, and Baja California workers regularly commuting to jobs in the United States. The analytical framework used was the staple theory developed largely in explaining the economic growth of Canada. The role of government policies (on both sides of the border), institutional arrangements, and the nature of the production functions for the exports are examined in an effort to determine the impact of the exports on economic activity.

PETER WEBB, Ph.D. Indiana 1969. Revolution and land reform: Mexico as a model for Latin America?

The thesis compares the roles that revolution and land reform can play in the economic development of the Latin American countries. As envisaged by the Alliance for Progress peaceful, democratic land reform can accompany economic development. On the other hand, the Mexican economist, Dr. Edmundo Flores advocates revolutionary Mexico as the model for underdeveloped Latin American countries. It is his thesis that violent revolution is a necessary condition for land reform in Latin America; peaceful and democratic political change is not possible. The thesis examines and evaluates the program of the Alliance and the proposition of Dr. Flores. Some counter-proposals are presented in conclusion.

THOMAS R. WEBB, Ph.D. Michigan 1969. A systems model for market development planning: Northeast Brazil.

This thesis presents a detailed description of a general systems model of a regional subsector of Northeast Brazil. The purpose of this modeling effort is to provide a framework for development planners to assess alternative investment programs in the coordination of market processes for an underdeveloped economy. Computer simulation and systems analysis are combined to build the model which details six production, distribution, and consumption sectors. Special "Planner Entry Routines" are developed to link the planner to the model structure.

JOHN F. WEEKS, Ph.D. Michigan 1969. Wage behavior, rural-urban income trends, and wage policy in Nigeria.

PHILLIP M. WEITZMAN, Ph.D. Michigan 1969. Planning consumption in the USSR.

CHARLES D. WHYTE, Ph.D. Ohio State 1969. Dynamic short-term estimating and forecasting models for wholesale beef price.

THOMAS F. WILSON, Ph.D. Columbia 1970. Labor force participation and business fluctuations: an analysis by cyclical stages.

FRANK ZAHN, Ph.D. California (Santa Barbara) 1969. A long-run growth model for Japan and a test of the low level equilibrium trap hypothesis.

A growth model for the Japanese economy was constructed and estimated to test the low level equilibrium trap hypothesis. The reduced form of the model, a difference equation exhibiting variable coefficients, was used to simulate the Japanese experience for the period, 1872 to 1936. An examination of the reduced form indicated that Japan did not experience the trap phase of development in the period. Thus, doubt is cast on the trap hypothesis for one of its most popular candidates, the Japanese economy.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

M. A. ABE, Ph.D. Wisconsin (Madison) 1970. Dynamic microeconomic models of production, investment, and technological change in the U.S. and Japanese iron and steel industries.

JOHN H. BERRY, Ph.D. Purdue 1969. A method for handling pecuniary externalities in relating firm and aggregate supply functions.

The objective of this study was to modify the firm assumptions used in previous studies to supply response in an attempt to obtain product supply functions that are more realistic. The model developed is a long-run area production adjustment model utilizing linear programming. The model permits interactions among farms, allows for pecuniary externalities on important inputs, accounts for changes in number of farms, and permits farm size to vary within the planning horizon of the model.

KUL B. BHATIA, Ph.D. Chicago 1969. Individuals' capital gains in the United States, an empirical study, 1947–64.

Realized gains are reported for tax purposes but the economic theorist would include accrued gains (whether realized or not) in income. A simple model for estimating accrued gains is developed. Capital gains accruing annually on corporate stock, real estate,

and livestock owned by individuals in the United States during 1947–64 are computed. The estimates alter both personal income and saving but the latter is affected much more than the official estimates of personal income.

FRANKLIN L. BROWN, Ph.D. Purdue 1969. Exact finite sample distributions of GCL structural estimators: a leading case.

WENDELL V. BROWN, Ph.D. Syracuse 1970. Empirical analyses of the secular distribution of income defined according to an accretion factor income concept.

GIORGIO CINGOLANI, Ph.D. California (Berkeley) 1970. Analysis of the dynamics of tree-fruit acreage in California's Central Valley.

The study is an empirical investigation of the acreage dynamics of orchard crops. Two quantitative techniques are tested and evaluated: an econometric time-series analysis of new plantings and removals of fruit trees, based upon the theory of investment behavior, within the framework of the neoclassical analysis of optimal capital accumulation; and a mathematical programming model applied to the allocation of land over time to orchard and field crops. The model is a multistage linear programming model combined with a generalization of the "recursive programming" idea of sequential decision processes. Both the econometric and the programming models are aggregate models and make use of secondary data, compiled from published and unpublished sources.

LAWRENCE P. COLE, Ph.D. Purdue 1969. A model of the U.S. economy, 1930–59.

This paper examines certain aspects of R. L. Basmann's simultaneous equations-model of the U.S. economy in which *GNP*, consumption (*GNP-GPDI*), *GPDI*, money supply, and interest rate are endogenous variables, and monetary base and an investment-shift variable are exogenous. The restrictions on the ranges of structural parameters are narrowed wherever possible so as to contrast the boundaries of the 21-dimensional structural parameter space. Results using constant dollar and current dollar data are compared, as are those using broad versus narrow definitions of the money supply. Also, various sub-periods were examined. The 1929–39 current dollar run gave the best results.

RONALD L. COOPER, Ph.D. California (Berkeley) 1969. The predictive performance of quarterly econometric models of the United States.

DONALD L. DEWING, Ph.D. Southern California 1969. Economic quality control.

Total quality costs which include the losses due to the production of discrepant materials and products and the costs involved in quality control. They are a significant part of product cost and are increasing rapidly with the increased demand for more complex and higher quality products. Using a literature search, economic analysis, and a case study, it is concluded that economic quality control systems reduce quality costs, improve product quality, increase revenue, reduce costs, and increase profits.

WILLIAM J. DUFFY, Ph.D. Pittsburgh 1969. OLS, 2SLS, LI/LGRV estimation of the time path of impact and delayed multipliers in a large econometric model, 1948–65.

A systematic investigation, using a quarterly econometric model, revealed that there have been significant structural shifts in a portion of the U.S. economy over the period 1948–65. A theoretical exploration of estimation problems connected with structural shifts was also initiated.

R. CLYDE GREER, Ph.D. Minnesota 1970. An analysis of price dispersion.

The total value of all resources expended in the procurement was determined for a set of manufactured dairy feed buyers. All buyers faced the same sources of supply. The dispersion among the set of effective prices was analyzed, and hypotheses explaining the dispersion tested. Variables considered were: purchase price exchanged between buyer and seller; additional cost of transportation; additional cost of farm; cost of knowledge; and cost (benefits) of the time relationship between procurement, feeding, and payment.

WILLIAM K. HALL, Michigan 1969. A queueing theoretic approach to the allocation and distribution of ambulances in an urban area.

ARCH W. HUNT III, Ph.D. Texas (Austin) 1970. Statistical evaluation and verification of digital simulation models through spectral analysis.

Fishman and Kiviat of the RAND Corporation have suggested the use of spectral analysis as a technique for evaluating the significance of statistics emanating from a simulation model. This author has developed this concept into a synthesized model and verified the model through a spectral analysis on data derived from the Logistics Composite Model (*LCOM*). *LCOM* was developed for the purpose of simulating aircraft flight and base support processes. Some forty statistics were evaluated through spectral analysis for confidence limits. Further, certain policy changes regarding inventory were implemented and a spectral analysis was performed to judge whether or not the separate results were statistically different.

YOUNG SIK JANG, Ph.D. State University of New York (Albany) 1969. A comparative study of interdependent and causal-chain systems in econometric model building.

HYMAN JOSEPH, Ph.D. Northwestern 1969. Costs of adjustment: an empirical study of business investment.

The adjustment process that is utilized to explain how a firm moves from its actual capital stock toward its desired capital stock is based upon economic variables. The speed of adjustment for each firm depends upon both the variables that affect the costs of adjustment and the relative gap between the desired capital stock and the actual capital stock.

JOHN H. KAGEL, Ph.D. Purdue 1970. Factor demand functions for labor and other inputs by northwestern wheat fallow farms.

JAMES L. KENKEL, Ph.D. Purdue 1969. An investigation of the exact finite sample distribution of the Von Neumann test statistic in a dynamical economic representation.

JAE WON LEE, Ph.D. New York 1969. Impact of technological change on the functional distribution of income.

STEPHEN D. LEWIS, Ph.D. California (Santa Barbara) 1969. Simulations of potential gross national product with a macroeconomic model.

The study contains a 12-equation non-linear model which is carefully predicated upon economic theory. The most interesting parts of the model include consideration of user costs, the desired stock of capital, and the utilization rate of capital enabling both static and dynamic simulations of actual and potential *GNP* to be carried out. The behavior of endogenous variables including the major components of *GNP*, employment, and labor force participation are examined.

RASIAH MAHALINGASIVAM, Ph.D. Toronto 1969. Market for Canadian refined copper: an econometric study.

The study is a contribution to the field of econometric research in the analysis of the demand for raw materials. It attempts, among others an assessment of the demand for Canadian copper by: derivation of a demand function using a two-level production function and estimation of a modified function incorporating the one way substitution effect of aluminum to copper; explaining the demand for imports of Canadian copper; analyzing the producers behavior when a commodity is one of the joint products; introducing the price determination behavior of the distributors of copper.

PARTHASARADHI MALLELA, Ph.D. Rochester 1970. Identification and estimation of large scale econometric models.

Necessary and sufficient conditions for the Standard Rank Condition to hold almost everywhere in the parameter space are derived. The nature of the Implied Restrictions on the Reduced Form is studied. Necessary and sufficient conditions for the identifiability of an equation of a model with lagged endogenous variables and autocorrelated disturbances are derived. A method of estimation of such models, with shortage of observations, is suggested.

MARVIN S. MARGOLIS, Ph.D. Purdue 1969. Testing hypothesis with confidence regions in simultaneous equations economic models.

A method for testing economic hypotheses in simultaneous equations models is presented. The testing method is restricted to nondynamical models and is valid for the general multivariate linear hypothesis and to a restricted extent for stochastic hypotheses. It tests structural hypotheses by expressing them as reduced-form hypotheses and uses confidence regions to test the reduced-form hypotheses. A geometrical example and two sample models illustrate the testing technique.

JAMES R. MARSDEN, Ph.D. Purdue 1970. A consideration of procedures for testing preferred economic hypotheses: derivation and application of procedures for testing preferred production functions.

This dissertation explores aspects of testing economic hypotheses with particular emphasis on the derivation of testing procedures in connection with proffered production functions. The first section deals with the presentation of the logical system and framework used throughout the paper. Testing procedures are then derived for use in testing proffered production functions. Two interpretations are outlined in detail, the testing procedures applied, and the results presented in the third section. The emphasis throughout is on an existence problem rather than an "estimating" problem.

MAHMOUD S. MARZOUK, Ph.D. Pennsylvania 1969. The predictability of predetermined variables in macro-econometric models.

JOHN M. MASON, JR., Ph.D. Michigan State 1970. An econometric analysis of some major exogenous determinants of national output.

There exist two extreme theories of national output determination. They are the Quantity Theory of Money and the Income Expenditure Theory. In this thesis an attempt is made to evaluate the two theories using a nine-equation econometric model. The model shows that although government expenditures have a greater initial impact on the level of aggregate activity, unadjusted for price changes, changes in the monetary base affect the economy over a longer period of time and exert a much greater accumulated effect.

BRIDGER M. MITCHELL, Ph.D. Massachusetts Institute of Technology 1969. Estimation of large econometric models.

MOSTAFA MOINI, Ph.D. Oklahoma 1970. Empirical verification in economics: classification and aggregation.

This study aims at evaluating the methodological significance of classification and aggregation in economics. It focuses on systems of Keynes and Leontief,

as specific media for the investigation of the more general methodological problems. A by-product of the study is the systematization of the literature on classification and aggregation in input-output analysis. Also, a disaggregative monetary model is formulated as a specific instance of applying the disaggregative strategy in theory formulation.

TIMOTHY D. MOUNT, Ph.D. California (Berkeley) 1970. An analysis of production behavior incorporating technical change.

Distinct production alternatives are identified by a set of Cobb-Douglas production functions. The analytical objective is to determine whether it is optimum to invest in capital embodying a new production method, or to expand existing facilities. Maximizing the present value of net returns to capital less discounted replacement costs is used as the decision criterion, subject to a constraint on investment expenditure. The model is applied to data for the dairy industry in California.

RODRIGO MUJICA, Ph.D. California (Berkeley) 1970. Satiation levels and consumer demand: analysis of a Chilean family expenditures survey.

TRACY W. MURRAY, Ph.D. Michigan State 1969. The impact of data errors on economic parameter estimation: a case study of international trade data.

The popular methods of estimating economic relationships where there exist errors in data make a set of assumptions regarding these errors. In this thesis, a method which requires two sets of observations on a given set of events is developed to test these assumptions. This method is applied to international commodity flows with the result that the usual assumptions are rejected. Alternative error probability distributions are specified and tested, and the results are applied to the problem of estimating international import demand elasticities.

FAROUK MUWAKKI, Ph.D. Nebraska 1969. Interindustry analysis of the economy of Syria for 1963.

PETER H. NIEHOFF, Ph.D. Nebraska 1969. Project timing.

QUAZI MD. M. RAHMAN, Ph.D. Texas A&M 1970. An optimal investment and financial control model: theoretical solutions and an application.

In this dissertation, a dynamic model for the firm was formulated as a discrete-time control problem and was transformed into a sequential linear programming problem. Interpretations of the theoretical solutions were made. The model was applied to farming on the Texas high plains where the water source is an irreplaceable underground aquifer.

POTLURI RAO, Ph.D. Chicago 1969. Some problems of estimation in cases of multiple production.

Multiple production is due to economic advantage, or technical restrictions. Econometric problems in each of these cases are different and are studied separately. Estimation of production functions and transformation curves under constant and variable elasticities are suggested. Several alternative test procedures for fixed proportions in products and factors are also suggested. Empirical examples are provided to illustrate these tests and estimation procedures.

LEROY H. ROCCA, Ph.D. California (Berkeley) 1969. Time-series analysis of commodity futures prices.

HIROSHI SADAMICHI, Ph.D. Johns Hopkins 1970. A synthesis of simultaneous equation estimators with an application to an inventory model.

The derived-form least-squares method is defined as the classical least-squares method applied to such a transform of a structural equation that the classical least-square assumptions may hold asymptotically. Most of available estimators are shown to be members of the derived-form least-squares family. A precise form of equation generating optimal inventory stock is derived from the assumption of profit maximization, and unitary elasticity of desired stock with respect to sale is found in many manufacturing industries.

WILLIAM E. SASSER, JR., Ph.D. Duke 1969. A finite-sample study of various simultaneous equation estimators.

GREGORY K. SCHOEPFLE, Ph.D. Purdue 1969. An application of exact finite sample distribution functions of structural coefficient estimators in economic statistical inference.

This dissertation is a qualitative and analytical study of finite sample distribution functions of GCL structural coefficient estimators and their application in the statistical inference associated with a simple four-equation non-dynamic simultaneous equations representation of the U.S. economy. Several small sample properties, including convergence to the parameter point and approximating distribution functions of the estimator distribution functions are discussed, and their application to the statistical inference of the structural coefficient estimates of the model under consideration is illustrated.

ANDREW D. SEIDEL, Ph.D. New York 1970. The formulation, estimation, and forecasting of nonstationary stochastic economic time-series.

HELIOS SILVESTRE, Ph.D. Cornell 1969. Demand analysis: An attempt to develop a methodology for detecting the points in time where structural changes took place.

A study of methods by which structural changes in demand can be identified and located in time. One method consisted of considering all possible divisions of the period of time covered by the data. A second method is called the smallest sum of sums squared residuals, and the third, the smallest sum of residual variances. The fourth method was based on construct-

ing equations for part of the sequence of data, then using this to predict forward and backward and testing for the significance of the difference between estimated and actual values. Several time-series were used for each method.

SATEESH K. SING, Ph.D. Harvard 1970. Investigations in production analysis.

DENNIS E. SMALLWOOD, Ph.D. Yale 1970. Estimation and prediction with the fixed coefficients, vintage model.

V. KERRY SMITH, Ph.D. Rutgers 1970. An economic evaluation of several econometric estimators of simultaneous equation systems.

This study reports the findings of research examining the impact of structure (sparseness and dimension of the endogenous variables coefficient matrix) upon the performance of several estimating techniques (OLS, 2SLS, LISE). Model I from Klein's *Economic Fluctuations in the United States*, the Klein-Goldberger model and the Dutta-Su Model of Puerto Rico were used in a Monte Carlo experiment. The results indicate that as models grow in size and relative sparseness OLS becomes the most desirable of the three estimators investigated.

VINCENT SU, Ph.D. Rutgers 1970. The specification, estimation, and evaluation of an economic relationship involving two independent distributed lags.

If an endogenous variable is conjectured to depend on two exogenous variables, then there could be two independent distributed lags involved in a single-equation model. If the distribution and the length of lags are known a priori, the unmeasurable structural model can be reduced to a measurable reduced-form model. Eighteen cases, each containing two differently distributed lags, were discussed. The specification of reduced forms and error terms and the estimation of parameters were thoroughly studied.

ANTONIO R. TEIXEIRA-FILHO, Ph.D. Purdue 1970. An evaluation of methodology employed in the estimation of farm level production functions.

This study dealt with an evaluation of the production function as applied to farm cross-sectional data. The following were determined or evaluated: the degree to which technical efficiency and price efficiency concepts were validly employed; the degree to which procedures employed in aggregation of input and output variables might have biased findings; the impact of competitive conditions on estimates of elasticities; alternative means of identifying production function parameters; the relative performance of ordinary least squares, non-linear least squares, and weighted least squares estimates.

J. SCOTT TURNER, Ph.D. Southern Methodist 1970. The measurement and analysis of production time as a factor of production in industrial processes.

This study investigates production time as an economic factor of production. Two basic models are developed in this study. The first is for the optimal utilization of production time, and the second is for the measurement of production time in an industrial process.

JOHN J. VALENTINI, Ph.D. California (Berkeley) 1969. Single family housing investment: a regional and national analysis.

FRITS VAN BEEK, Ph.D. Maryland 1970. An econometric model of international trade in machinery and equipment.

MARK A. WALKER, Ph.D. Purdue 1970. Messages and iteration of joint decision-making mechanism in finite environments.

CARL R. WENDOLOSKI, Ph.D. Harvard 1970. Personal services in the United States: an analysis of employment and expenditures.

The thesis presents an analysis of differences in estimates of the income elasticity of demand for personal services which are obtained when different types of data are used. Also, an attempt is made to account for areal differences in expenditures and employment. In addition, secular trends in employment and expenditures in the United States are traced and analyzed.

JOHN C. WIGINTON, Ph.D. Carnegie-Mellon 1970. Discrimination among economic models: a Bayesian analysis with application to aggregate demand for money.

NORMAN K. WOMER, Ph.D. Pennsylvania 1970. Estimating the relations of an input-output model with variable coefficients.

KAN HUA YOUNG, Ph.D. Columbia 1969. Demand for sugar in the United States: a synthesis of time-series and cross-section analyses.

A general time-series model which incorporates the permanent income and partial adjustment hypotheses and a general cross-section model which incorporates the relative income and demonstration effect hypotheses are investigated. Furthermore, a more general model which combines these two models is also constructed. Through this general time-series and cross-section model, an appropriate procedure for pooling (macro) time-series and (micro) cross-section data is developed. These models are employed for empirical analyses.

Monetary and Fiscal Theory, Policy and Institutions

MOTOO ABE, Ph.D. Pennsylvania 1969. A monetary model of the Japanese economy.

The thesis treats a monetary model of the Japanese economy, using quarterly data from 1952–I to 1964–IV. In the model the author attempts to analyze the interactions of monetary and non-monetary

factors in the economy and to measure the effects of monetary and fiscal policies. The model contains twenty equations and uses more monetary variables than other models of this size. The system is non-linear and is characterized by sixteen simultaneously determined variables.

POLLY R. ALLEN, Ph.D. Brown 1970. Money and growth in open economies: a neoclassical approach.

The influence of the growth rate of money on long-run capital intensity is examined in a neoclassical model of open economies under flexible exchange rates. Increased money growth raises capital intensity in countries that produce the capital good or import capital and face relatively elastic demand for exports, similar to conclusions for a closed economy. But for the capital-importing country facing relatively inelastic world demand for exports, increased money growth lowers capital intensity.

ROBERT N. ANDERSON, Ph.D. Colorado State 1969. A systems analysis of public welfare.

The problem of public welfare in Colorado is defined. Details and implications of various alternative public welfare programs are examined, with particular emphasis on the proposed state "takeover" of public welfare programs in terms of administration and financing. Arguments are examined as to the merits of permitting the countries to continue in their functions within the present system. Abstract models are presented for evaluating and planning welfare programs.

DAVID AULT, Ph.D. Illinois (Urbana) 1969. The determinants of world steel exports: an empirical study.

ERNST BALTENSPERGER, Ph.D. Johns Hopkins 1969. Economies of scale, firm size and concentration in banking.

The thesis deals with uncertainty, economies of scale, and their role in explaining firm sizes and concentration in banking. A first part discusses the mechanisms by which uncertainty can create scale economies in banking. In a second part, some phenomena of bank size and concentration are studied for which the described scale economies can offer an explanation, and which, in turn, can give us some confirmation about the existence and significance of these economies.

STEPHEN BARNETT, Ph.D. New York 1969. The household demand for demand deposits, savings and time deposits at commercial banks, and savings and loan shares.

ROBERT J. BARRO, Ph.D. Harvard 1970. Inflation, the payments period, and the demand for money.

Chapter I develops a theory of microeconomic response to inflation. Changes in such key decision variables as the payments period and the fraction of "monetized" transactions are the vehicles by which individuals reduce average real cash holdings in response to higher rates of inflation. When inflation rates vary over time, the optimal response depends on a distributed-lag of actual inflation rates, according to a variable coefficient-of-adjustment. Chapter II applies the theory to an empirical study of demand for money in several cases of hyperinflation. The results provide empirical support for the model. Chapter III describes some extension of the model to the areas of inflationary finance and the welfare cost of inflation.

ROBERT A. BARRY, Ph.D. Yale 1970. Indicators of monetary policy as intended by the federal open market committee, 1953–60.

JAMES T. BENNETT, Ph.D. Case Western Reserve 1970. A spectral analysis of cyclical fluctuations in money and business.

The dissertation examines Friedman's monetary theory of the business cycle with spectral techniques. In contrast to Friedman's findings, the estimates indicate that, in the business cycle, business activity leads changes in the money stock. The changes are not highly correlated, and fluctuations in business are approximately equal to the magnitude of fluctuations in money. The money stock is strangely influenced by a feedback mechanism; therefore, the money supply is an unreliable indicator of monetary policy.

JULIO BERLINSKI, Ph.D. Harvard 1970. The behavior of provincial finances in Argentina.

LLOYD C. BILLINGS, Ph.D. Arizona 1969. Determinants of member-bank borrowing from the 12th Federal Reserve Bank, 1953–67.

DAVID E. BLACK, Ph.D. Massachusetts Institute of Technology 1969. Inequalities in effective property tax rates: a statistical study of the city of Boston.

An analysis of residential effective property tax rates reveals systematic variations of significant magnitude. Rates tend to be related positively to the number of families per structure; positively to the neighborhood density of low quality housing and nonwhite families; and negatively to neighborhood family income. One of the most important of several suggested explanations for these variations is the failure to adjust assessments to changing market values. Estimates of effective tax rates are based on 20,000 observations of residential property transactions.

RICHARD E. BOND, Ph.D. Maryland 1970. Deposit composition and commercial bank earnings.

DONALD BOOTH, Ph.D. California (Los Angeles). An analysis of private land use controls and private cities as systems to produce public goods.

There are differences between public and private cities in the recognition of externalities based upon cost and benefit calculations as perceived by decision makers. These differences result from the nature and distribution of property rights. Sectors of existing public cities and complete private cities need externalities through land use regulations either by the disag-

gregation of property rights or the dispersion of property rights.

JAMES M. BOUGHTON, Ph.D. Duke 1970. The effect of an active market in federal funds on the transmission of monetary policy.

GEORGE BUDZEIKA, Ph.D. New York 1970. Lending to business by New York City banks.

GEORGE E. CARTER, Ph.D. Clark 1970. Canadian conditional grants postwar.

A theory of federal conditional grants is developed, and their use in Canadian federalism postwar is appraised. An attempt is made to measure income redistribution effected by these transfers. Provincial dissatisfaction with the design and other features of the grant programs is examined along with the recent federal offer to substitute federal tax abatements of personal income tax for the three major grants.

VICTOR P. CARTER, Ph.D. Case Western Reserve 1970. The postwar growth in the flow of funds from the Canal Zone to the Republic of Panama.

From 1946 to 1966 the annual flow of funds from the Canal Zone to the Republic of Panama grew from $50,551,000 to $106,742,000. After presenting the background of the Canal-Panama relationship, the study undertook to analyze the growth in the flow of funds to ascertain the quantitative effect of the following four factors: (1) growth in international trade and transportation activities of the Canal, (2) expansion of the U.S. economy, (3) enhancement of Panama's ability to supply the goods and services purchased by economic units within the Canal Zone, and (4) changes occasioned by political events.

SANDRA B. COHAN, Ph.D. Michigan 1969. The determinants of supply and demand for certificates of deposit.

WILLIAM B. CONWAY, Ph.D. Minnesota 1969. Capitalization of the property tax: an empirical study.

This thesis includes a critical review of previous statistical studies of property tax capitalization and an attempt to measure the extent of such capitalization on residential housing in an urban area. Data was gathered in the Minneapolis-St. Paul metropolitan area. The results in general did not support the capitalization hypothesis, and suggest that the inequalities in the administration of the property tax are exaggerated.

JOHN E. COOPER, Ph.D. Washington State 1970. The impact of proposed negative income tax plans on the economy of the state of Idaho.

SUZANNE CUTLER, Ph.D. New York 1970. An investigation of the dual banking system: the example of New York state commercial banking.

This study investigates the influence of the dual banking system on the operation of the banking industry from 1950 through 1967, using the example of New York state commercial banking. The analysis of selected regulatory and supervisory functions offered few defenses for continuation of the dual system and indicated the possibility of major impairment to the dual system particularly because of the increased defections from the state systems.

LAWRENCE DEWITT, Ph.D. Syracuse 1969. The inflation-unemployment trade-off: the distribution of benefits and burdens.

MICHAEL G. DWORKIN, Ph.D. Michigan 1969. Expectations and the term structure of interest rates.

THOMAS W. EPPS, Ph.D. Duke 1969. Financial asset prices: A theoretical and empirical study.

STANLEY FISCHER, Ph.D. Massachusetts Institute of Technology 1969. Essays on assets and contingent commodities.

DONALD R. FRASER, Ph.D. Arizona 1969. Commercial bank market structure and farm loan rates: the evidence from Texas.

JOEL S. FRIED, Ph.D. Northwestern 1969. A generalized Markowitz model of bank portfolio.

The dissertation develops and tests a Markowitz type model of bank portfolio selection that includes probabilistic liquidity constraints. Actual bank portfolios did not conform to the risk and liquidity constrained portfolios generated by the model. However, using *ex post* returns it was shown that for given risk and liquidity levels the generated bank portfolios did statistically better than actual portfolios suggesting that present portfolio selection procedures are inefficient.

HAROLD N. FRIEDMAN, Ph.D. Alabama 1969. John Kenneth Galbraith's contributions in the areas of fiscal policy, inflation theory, and social balance.

The study hypothesizes that Dr. Galbraith's ideas have appreciable present and future applicability, and as such, are worthy of consideration. The purposes of the study are to assess Galbraith's contributions in the areas of fiscal policy, inflation, and social balance, and to make available to the reader specific theoretical formulations, federal policy history, and certain Galbraithian concepts in the aforementioned areas.

SHIGEYUKI FUKASAWA, Ph.D. Columbia 1970. A variable lag pattern in the formation of expected price changes.

This dissertation concerns the determination of the length of lag in inflation expectations. Two kinds of hypotheses are discussed: One assumes the length of lag in expectation to be negatively correlated with the actual or expected intensity of inflation, and the other assumes it to be negatively correlated with variability of past price changes. Empirical results using data from European hyper-inflations and the U.S. postwar

period showed conclusive evidence for the latter hypothesis.

ISMAIL A. GHAZALAH, Ph.D. California (Berkeley) 1969. U.S. government aid under P.L. 874 to federally impacted school districts.

RONALD D. GILBERT, Ph.D. Oklahoma State. An econometric analysis of the Friedman-Cagan money supply hypothesis.

The recent literature in monetary economics reveals a revived interest in money supply theory. The present study involves the specification and estimation of a money supply model which uses the Cagan money supply identity. The model, which includes eight equations and three identities, not only includes the money supply as an endogenous variable, it also includes the reserve-deposit and currency-money ratios as simultaneously determined endogenous variables. The parameters are estimated for 1934–67, and for 1952–67.

ROY F. GILBERT, Ph.D. Michigan State 1969. The demand for money: an analysis of specification error.

Empirical results reported in the literature support a large number of conflicting theoretical hypotheses concerning the "correct" specification of the demand for money function. Statistical tests for the presence of specification error were applied to two hundred different money functions using annual data for the period 1914–58. The test results indicate that most of the empirical money functions reported in the literature since 1955 are *significantly* misspecified.

ERROL GLUSTOFF, Ph.D. Stanford 1969. Studies in monetary theory and consumer behavior.

JEAN M. GRAY, Ph.D. California (Berkeley) 1969. The term structure of interest rates in the United States: 1884.

KENNETH V. GREENE, Ph.D. Virginia 1968. Tax institutions, income distributions, and the transfer of functions in a federalism.

MICHAEL G. HADJIMICHALAKIS, Ph.D. Rochester 1970. Money in the theory of economic growth.

The postulated market clearance in equilibrium models is examined through the introduction of price-adjustment mechanisms and expectations. Several models in the literature are found inconsistent. The inconsistency is not remedied through the process of capital acculmulation. Conditions for short-run consistency and long-run convergence are given. When the transactions aspect predominates, a second equilibrium arises. Making savings dependent on expectations, a synthesis of Solow-Tobin models is achieved, with the examined models as special cases.

EILA HANNI, Ph.D. Yale 1970. Inflation in postwar Finland.

PAUL M. HAYASHI, Ph.D. Southern Methodist 1969. Demand for unborrowed excess reserves by commercial banks.

The purpose of this study is to develop a theory of UER (unborrowed excess reserves) as an application of inventory models under uncertainty. The theory of the demand for UER developed in this study suggests that desired UER are a function of the required reserve ratio, total deposits, the expected reserve loss per dollar, the standard deviation of the reserve loss distribution, and a bank's aversion to risk, which in turn is a function of the interest rate, the cost of borrowing reserves, the cost of reserve management and the expected reserve loss per dollar.

WILLIAM L. HENNING, Ph.D. Ohio State 1969. An economic analysis of savings and income in rural America.

ROGER H. HINDERLITER, Ph.D. Washington (St. Louis). An econometric investigation of short-run reserve position adjustments of New York City banks: 1962–67.

JOSEPH F. HUMPHREY, Ph.D. Southern California 1969. The role of the manager of the Federal Open Market Committee: 1951–61.

Evidence gathered from the minutes of the Federal Open Market Committee and other publications of the Federal Reserve System indicates that the manager enjoyed considerable independence in his actions. This evidence has two implications: the importance of the manager has been underestimated and may provide an explanation of some of the as yet unresolved issues concerning the Federal Reserve; one individual, rather than the duly constituted body, should not be exercising this power.

THOMAS M. HUMPHREY, Ph.D. Tulane 1970. The development of the quantity theory of money in the United States 1880–1930.

The quantity theory is traced through the Bimetallism Controversy, Newcomb-Fisher era, and 1920's stabilization debates and business cycle research. The study shows that earlier theorists: developed a subtle, more complete monetary theory distinct from the crude version employed in their empirical work; advocated monetary reform because (contrary to the "neutrality" implications of their crude model) they feared real-sector disruptions caused by an erratically changing money supply; and presaged concepts and characteristics prominent in the modern Chicago approach.

DAVID N. HYMAN, Ph.D. Princeton 1969. A behavioral model for commercial banking.

A two-asset, one-liability model of commercial bank behavior is constructed. The model determines optimal portfolio allocation *and* size for a utility maximizing bank, operating under conditions of perfect competition in all markets. The comparative static properties of the model explain some of the interde-

pendencies that exist between the asset and liability sides of the balance sheet. The results for the perfectly competitive model are qualified by introducing imperfect competition into the bank's markets.

DUDLEY D. JOHNSON, Ph.D. Virginia 1968. The pure theory of trade credit and its influence on monetary policy.

DAVID M. JONES, Ph.D. Pennsylvania 1969. Member bank borrowings from the Federal reserve: an analysis.

Member bank borrowings from the Federal Reserve are found to be responsive to interest rates, given bank reserve needs. The results show that the discount rate-Treasury Bill rate spread, which had the dominant influence on borrowings in the 1950's, was replaced in the 1960's by the dominance of the discount rate-Federal funds rate spread. The findings also suggest that banks complete most of the adjustment of actual borrowings to "desired" levels within a month.

GREGORY V. JUMP, Ph.D. Michigan 1969. An econometric model of the financial sector of the U.S. economy.

WILLIAM M. KEMPEY, Ph.D. New York 1970. Is monetary policy effective? An analysis of the efficacy of monetary policy on the basis of a multi-sector econometric model for the United States: 1940–66.

GEORGE K. KEYT, Ph.D. Minnesota 1970. Price level variability and the demand for money.

Portfolio composition implications for wealth maximization models are investigated under various assumptions about present asset prices, interest rates, and price expectations. The choice is made between money and non-monetary assets, both of which have variable expected real values, by an individual whose utility function is assumed to be quadratic. In general, it is found that making the real value of money variable modifies the portfolio asset composition response to changes in other economic variables.

HI KYUNG KIM, Ph.D. Houston 1969. The Gurley-Shaw hypothesis reexamined: A disaggregated approach.

This dissertation tests the relationship between the proportion of a given financial asset (money and several types of near monies) to total financial assets of nonfinancial economic units (households and nonfinancial businesses) to the prevailing relative yields among different types of financial assets and the current monetary policy. The main conclusion is that the proportion of liquid assets to total financial assets is a stable function of the difference in yields among financial assets.

BRUCE W. KIMZEY, Ph.D. Washington State 1970. Local government finance and the future of the Seattle area.

JAMES V. KOCH, Ph.D. Northwestern 1969. The demand functions of the household sector for financial assets: an econometric study.

This dissertation is a theoretical and empirical study of the financial asset demand functions of the household sector. The theoretical work concentrated upon effects of different types of assumptions upon the predictions of a utility maximization model with a wealth constraint. The empirical work concentrated upon the lag structure of the demand functions and upon the question of the homogeneity of the demand functions.

ALAN D. KRAUS, Ph.D. Cornell 1969. The forecasting accuracy of models of the term structure of interest rates.

ATIF A. KUBURSI, Ph.D. Purdue 1969. The financial sector and the economy: impact and role under alternative hypotheses.

This dissertation focuses on the failure of current macroeconomics to present a meaningful treatment of the financial sector of an economy whether in social accounting systems or in macroeconomic hypotheses associated with such systems. Consequently a new convention is adopted which allots to the financial sector a separate and distinct stature, and macro-models are altered to introduce new variables which are unjustifiably neglected in the current literature on macro-theory.

JAY J. LADIN, Ph.D. Purdue 1969. Factors associated with variations in state and local government efforts to provide public goods and services.

JERRY W. LEE, Ph.D. Florida 1969. Changing liquidity patterns in the American banking system.

DAVID LOY, Ph.D. Kansas 1970. An allocation formula for federal grants to state public higher education.

ALAN S. MCCALL, Ph.D. Kentucky 1969. A statistical study of market structure and performance: small and medium-sized commercial banks.

STUART MCFADYEN, Ph.D. California (Berkeley) 1969. Home-owner grants; the British Columbia experience.

DENNIS J. MAHAR, Ph.D. Florida 1970. Federal-state fiscal relations in Brazil.

JOHN D. MANGOLETSIS, Ph.D. Rice 1970. The structure of assets and liabilities of the financial intermediaries in Brazil, 1952–65.

The dissertation presents the sectoral liquidity structure of the assets and liabilities of the financial intermediaries in Brazil at selected dates of the period 1952–65 and seeks to develop a method for the study of financial intermediation using the data for Brazil at the selected dates. The concept of financial structure, the structure of assets and liabilities or uses and

sources of funds, is complemented by two concepts, viz., the coefficient of financial transfer and the financial transfer structure which, unlike financial structure, apply exclusively to financial intermediaries. The theoretical foundations of the method developed in the dissertation are due especially to Professors J. G. Gurley and E. S. Shaw and Professor R. W. Goldsmith.

RICHARD G. MARCIS, Ph.D Kentucky 1969. Commercial bank activity in the municipal bond market: 1947–1967.

This study focuses on the determinants of commercial bank purchases and sales of municipal bonds for the period from 1947 to 1967. The specific hypothesis examined is that commercial banks buy and sell municipal bonds in response to changing loan demands and credit conditions. Multiple regression analysis reveals that this hypothesis is substantiated only for the larger money market banks located in New York City.

Michel I. Marto, Ph.D. Southern California 1969. A money supply model for Jordan.

An empirical money supply function is estimated for Jordan by developing a simultaneous equation econometric model and solving it to obtain a reduced form equation for the money supply. Data for the period 1951–66 were used to fit the model. Statistical tests indicate that the money supply is satisfactorily explained as a function of lagged foreign assets, exports, transfers, local assets of the Central Bank, and the demand and time deposits of the Treasury.

LYNN C. MAXWELL, Ph.D. Nebraska 1969. The rate of return for schooling at the University of Nebraska.

HERBERT MAYO, Ph.D. Rutgers 1970. Theoretical linkages between the real and monetary sectors.

The dissertation considers various means for transferring monetary change to the real markets. The content of the paper falls into two sections: a discussion of the post General Theory literature and a discussion and extension of the linkages suggested by Keynes in his *Treatise on Money*.

PETER B. MEYER, Ph.D. Wisconsin (Madison) 1970. Citizen response to personal income taxation in low income countries: the application of household survey data to the analysis of taxation (a case study of the Philippines).

Data: 1965 Philippine Bureau of Census and Statistics Survey of Households, 4800 observations with tax paid and derivable tax due. Method: distinct analyses of base, rates, and administration of tax to yield recommendations for change; estimated determinants of tax payment and compliance. Findings: excessive exemptions, low initial rates; 14 percent overall compliance; withholding (good private, poor government) major revenue source; growth will increase yields, also relative burden on wage earners. Need new taxation at source instruments.

MARK J. NELSON, D.B.A. Oregon 1969. Time and savings deposit growth and small bank performance in the West, 1961–64.

The objectives of this study were: 1) to examine the relationship between small western bank holdings of time and savings deposits, the rates paid on these deposits, and the banks profitability as measured by return on assets and return on capital; and 2) to examine the relationship between small bank holdings of time and savings deposits, the rates paid on these deposits, and the banks profitability as measured by by average yield on loans, service charge rates, and effective rates paid on time and savings deposits.

WALTER E. NICHOLSON III, Ph.D. Massachusetts Institute of Technology 1969. A macroeconomic study of household asset choice.

This thesis examines the aggregate savings behavior of the Household Sector of the U.S. economy over the period 1952–65. Quarterly Flow of Funds Accounts data are used to investigate several hypotheses about household behavior. The major questions considered include: the household budget constraint and the role of "buffer stock assets"; interrelationships between credit and asset demands; substitutability and complementarity between various assets; and specification of "the" interest elasticity of consumption. While the econometric results are not consistently good, the Flow of Funds data do seem to be more amenable to investigation than would have been expected a priori.

DONALD NIELSEN, Ph.D. Syracuse 1970. An analysis of selected determinants of inter-county taxable property resources differences in New York State 1955 to 1966.

SAYED M. NIMEIRI, Ph.D. Syracuse 1970. Taxation in a developing country: a case study of the tax system of the Sudan.

CHONG KEE PARK, Ph.D. George Washington 1970. An analysis of income elasticity of the tax yield in Korea.

This study attempts a quantitative analysis of Korea's tax behavior as a function of national income and suggests income elasticity of yield as an important criterion for evaluating tax policy. Revenue yields from income taxes are highly elastic with respect to *GNP*. Although the overall income elasticity of commodity taxes is only 0.8, yields from certain commodity items, e.g., paper, cement, plywood, plate glass, are nevertheless highly responsive to changes in *GNP*.

YUNG CHUL PARK, Ph.D. Minnesota 1969. Asset preferences and the effects of monetary and fiscal policy in alternative static macroeconomic models of income determination.

The study analyzes the effects of monetary and fiscal controls, taking into account substitution and complementarity relationships among various assets, in three different models: a purely financial model, a neoclassical model, and a neo-Keynesian model.

ROY L. PEARSON, Ph.D. Virginia 1968. The equilibrium income tax structure: analysis in a democratic decision model.

ERIC E. PEDERSEN, Ph.D. New York 1970. A study of Swedish fiscal policy from 1950 to 1964.

This dissertation is an evaluation of the nature and effectiveness of Swedish fiscal policy, as seen in the context of Swedish institutions, social policy, and political goals.

DOUGLAS D. PETERS, Ph.D. Pennsylvania 1969. The stability of bank deposits.

In the postwar years bank loan expansion could be achieved through sale of Treasury securities, and it was thought that bank deposits might no longer behave in the same unstable manner as they had prior to 1930. Several models of bank lending were estimated which contrasted the postwar and earlier years. The results confirmed the changed structure of bank lending. Deposits were more stable, but questions remain whether the banking system contributed to economic stability.

DONALD PHARES, Ph.D. Syracuse 1970. Analysis of state-local tax burdens.

JULIUS C. POINDEXTER, JR., Ph.D. North Carolina (Chapel Hill) 1969. Asset shifts involving currency and the strength of monetary controls.

The thesis analyzes the impact on the strength of monetary policy of the existence of systematic shifts in currency demand in response to changes in an array of variables (constraints, opportunity costs, etc.) usually assumed to affect demand for other financial assets but ignored as a determinant of currency demand in traditional credit expansion models. The analysis employs both partial equilibrium techniques and a simultaneous equation macroeconomic model. Econometric evidence is drawn from existing large-scale models and from the author's own OLS regressions.

RONALD B. PRUET, Ph.D. Texas (Austin) 1969. Post World War II debt management in the United States.

Public debt management in its narrowest sense involves choosing the terms and maturities for government securities which are issued in Treasury financing operations. This seemingly simple task, however, may affect both liquidity and the interest rate structure in the national economy and deserves continuing examination. The dissertation examines three theoretical alternative approaches to debt management; countercyclical, procyclical, and neutral. Next the actual financing operations of the Treasury from March 1951 to July 1967 are reported. The major purpose of this study is to evaluate these Treasury financing operations in the light of the three alternative approaches.

ALFRED R. RECTOR, D.B.A. George Washington 1970. The compatibility between the inheritance tax law and the property law in the District of Columbia.

This dissertation examines the relationship of death tax administration in the District of Columbia to the applicable property law on which such tax purports to be based. Although there are extra inequities and administrative burdens in the inheritance tax, its economic incidence in the aggregate is not substantially different from that of an estate tax. Therefore, an estate tax should be substituted for the District's present inheritance tax.

GEORGE P. RONIGER, Ph.D. Columbia 1970. A general grant for the States: A consideration of its objectives, justification, and effects.

The probable effects of Federal revenue sharing are considered in light of its objectives and justification. A distribution of funds not based on state tax effort would not reduce the price of public services and would not overcome existing restraints to expenditures. Revenue sharing with an effort formula would reduce the price of public services and would result in small additions to state expenditures. The major portion of shared funds would substitute for state taxation.

PETER S. ROSE, Ph.D. Arizona 1969. Debt management: a comparison study of Bills Only and advance refunding, 1953–65.

ROBERT SCHROCK, Ph.D. Kansas 1970. The economics of commercial bank debentures.

FREDERICK D. SEBOLD, Ph.D. Boston College 1969. An investigation of the short-run shifting of the corporation income tax.

The dissertation uses both theoretical and empirical tools to analyze firms' reactions to variations in the corporation income tax rate. A utility-maximization model is developed to explain the potential rationale behind shifting. Then, a simultaneous-equation econometric model is specified and estimated to ascertain the degree of shifting taking place in the corporate manufacturing sector. The data are consistent with the view that the burden of the tax is borne fully by corporations.

ABDELALEEM M. A. SHARSHAR, Ph.D. George Washington 1970. A comparative study of theories of public enterprise.

Public ownership subjects the public enterprise to multiple criteria of performance. This dissertation examines the consistency of the major elements of such criteria, i.e., growth, efficiency, and social welfare in USSR, U.K., and India. Graphic models of public enterprise are built for each country. The pur-

suit of conflicting objectives causes external diseconomies, recognition of which reduces the attractiveness of the external economies of public undertaking in developing countries. Alternative performance criteria are developed in this study.

DAVID K. SHEPPARD, Ph.D. Harvard 1970. The growth and role of U.K. financial institutions, 1880 to 1962.

This is an empirical examination of the annual balance sheets of leading U.K. financial institutions from 1880 to 1962. The first part of the work has a descriptive orientation. The growth and the change in the market shares of various institutions' total liabilities, of the money stock, of certain money substitutes, and of certain forms of financial institutions' assets are represented. In the second part, the data is processed so as to show how well such financial variables as money, encashable assets, interest rates, and measures of credit availability performed as determinants of various forms of private expenditure.

ROGER E. SHIELDS, Ph.D. Virginia 1969. Economic growth with price deflation, 1873–96.

CHARLES T. SMITH, Ph.D. State University of New York (Buffalo) 1970. Demand for short-term government debt.

JAMES F. SMITH, Ph.D. Southern Methodist 1970. Determinants of the volume of consumer installment credit since 1948.

This is an econometric study, utilizing the generalized stock adjustment model of Zellner, of the determinants of the four categories of consumer installment credit, namely automobile credit, other consumer goods credit, repair and modernization credit, and personal loans.

ROGER S. SMITH, Ph.D. California (Berkeley) 1969. A study of local income taxes: with special reference to the city and county of San Francisco.

THEODORE R. SMITH, Ph.D. Claremont School 1969. A comparative analysis of real property and land value taxation.

The theoretical implications of land value and real property taxation are systematically analyzed through the application of relevant economic theory, including the development of a locational model. A second phase of the study involved testing the theoretical findings through the use of a case study. A land value tax is simulated for the city of San Bernardino, California, in an attempt to isolate the relative differences between the two taxes under consideration.

ROGER W. SPENCER, Ph.D. Virginia 1969. Adjustments of selected markets in tight money periods.

GEORGE F. STARNER, Ph.D. Ohio State 1970. Variation in the level of taxes on property: an Ohio case.

RICHARD T. STILLSON, Ph.D. Stanford 1970. An analysis of financial intermediation in the context of financial development.

A model of a financial firm is used to derive demand and supply functions for financial securities. It is shown that financial innovations shift these functions and result in decreasing the interest rate differential between loan rates and borrowing rates. Two case studies show the significance of these effects. Financial innovations in Japan (1868–1914) tightened the rural interest rate structure. Financial services provided in Malaya (1905–14) affected plantation and small holding investment in rubber.

THOMAS M. SUPEL, Ph.D. Minnesota 1969. A two-period balance sheet model for banks.

This dissertation is basically an extension of Tobin's work on financial models which considers the effect of liability characteristics on asset management. An explicit profit maximizing strategy is developed for portfolio adjustments at the beginning of the second period. The model then permits an analysis of the portfolio impact of both asset and liability uncertainty as well as a comparison of two-period versus myopic investors.

CRAIG E. SWAN, Ph.D. Yale 1970. The behavior of financial institutions: an econometric analysis with special attention to mortgage markets and residential construction.

JULIAN H. TAYLOR, Ph.D. Iowa State 1969. Wartime finance and the term structure of interest rates.

This dissertation presents an alternative test of a model presented by Modigliani and Sutch, and it investigates the previously uninvestigated financial developments which occurred during World War II. An alternative theory of the term structure of interest rates is developed which accords a much larger role to the horizon than is allowed by traditional expectations models. This theory is then tested with wartime data which is selected because of the vigorous supply activity. The results support the theory which combines both the expectations hypothesis and the segmentation hypothesis.

WILLIAM T. TERRELL, Ph.D. Vanderbilt 1970. The term structure of interest rates, portfolio theory, and the role of length to maturity in selecting U.S. government securities.

WILLIAM H. TOTTEN, Ph.D. Southern California 1969. The effect of wealth on consumption: empirical tests for significance and lags.

This dissertation examines the method in which the wealth effect has been tested in empirical models and concludes that existing tests have been carried out under widely differing circumstances and hence cannot be considered as collective evidence. New tests are set up to examine the sensitivity of wealth effects to different specifications with this same data. The results generally support the hypotheses of a strong

wealth effect. The results are particularly sensitive to the degree of autocorrelation in the residuals. When autocorrelation is reduced or removed the results provide stronger support for the hypothesis.

RAUL TREVINO-WESTENDARP, Ph.D. Columbia 1970. On the distribution of free governmental services: police protection in New York City.

ANTHONY M. TUBEROSE, Ph.D. Texas (Austin) 1970. The effects of changes in margin requirements.

The imposition of margin requirements (the maximum percentage of stock purchases which can be transacted on borrowed money) is a selective control influencing the flow of credit into the stock market. This dissertation examines the implications of this ability to vary the proportion of stock purchases transacted with borrowed money. Through the use of regression analysis changes that occur over the 90 day period prior to a margin change date are compared to the changes that occur 90 days subsequent to the change.

DONALD P. TUCKER, Ph.D. Massachusetts Institute of Technology 1969. Adjustment lags in macroeconomic systems and their implications for monetary policy dynamics.

JAMES R. UKOCKIS, Ph.D. Indiana 1970. The fiscal impact of federal aid on state and local governments: a stimulant or a substitute.

The study attempts to answer the question posed in the title by developing a number of alternative definitions of stimulation and substitution, then empirically tests for their existence using mainly U.S. Bureau of Census Data.

RAYMOND C. VARS, JR., Ph.D. California (Berkeley) 1969. The tax cost of freeways in San Francisco.

DONALD F. VITALIANO, Ph.D. City University of New York 1969. Impact of the interest on federal debt upon the distribution of income.

CHARLES WALDAUER, Ph.D. Syracuse 1969. The fiscal impacts of intergovernmental aid on local governments in Onondaga County, New York.

The major empirical findings are: special purpose aids do not cause local funds to be diverted away from unaided functions; school aids tend to substitute for local taxes in financing education; only town highway maintenance expenditures are stimulated by aid; and local governments sharing the same tax bases are fiscally interdependent, as the impacts of aid are transmitted to other governments via changes in the fiscal pressures on these tax bases.

CARLA M. WARBERG, Ph.D. Wisconsin (Madison) 1970. Evaluation of the short-run monetary policy system of the United States, 1951–68.

The purpose of this study is to establish a framework for the evaluation of short-run monetary policy systems as measured by the behavior of the money supply; then using quarterly data, to evaluate the U.S. system from 1951 through 1968. A second section defines a negative feedback control system for the monetary policy system and evaluates the behavior of the actual system with respect to this desired behavior. A third section tests some single-equation short-run models of income determination. A section using historical data is also included.

JAMES H. WARREN, Ph.D. Oklahoma State. An economic model of local education expenditures within an individual utility maximization framework: a theoretical and empirical examination.

In economic models, government spending is usually considered to be a "given," predetermined by the political process. A premise of this investigation is that such a process must ultimately be derived from individual choice. It is further assumed that individuals evaluate fiscal programs in terms of their own utility functions. An economic model of local public education spending is developed, within the individual utility maximization framework, and empirically tested. County areas in Virginia and South Carolina are used in the empirical test. A form of multiple regression analysis, using data for 1962, is employed to assess the explanatory potential of the theoretically constructed economic model.

ALEX R. WEAVER, Ph.D. California (Berkeley) 1969. A study of term structure theories.

HARLAND W. WHITMORE, Ph.D. Michigan State 1969. Empirical testing of the Friedman-Meiselman hypothesis.

The Friedman-Meiselman analysis is extended using revised data and alternative definitions of autonomous expenditures and the money supply. The major extension of the FM analysis is the examination of the dynamic properties of an aggregate econometric model of the U.S. economy. Primary emphasis is placed on an investigation into the relative effectiveness of the money stock and government expenditures to influence real net national product.

JOSEPH R. ZECHER, Ph.D. Ohio State 1969. An evaluation of four econometric models of the financial sector.

This research explores the logical content of four econometric models of the financial sector. It simulates the policy implications of the models, and reviews and analyzes such characteristics of the models as functional forms, data used, and linkage hypotheses. Some of the more important common characteristics are that the monetary base, the two required reserve ratios, and the discount rate reduce to one, or at most two independent policy instruments; and that none of these models are consistent with the hypothesis that accelerations in income or shifts in the economic structure over the business cycle dominate the level or rate of change of the money stock. Major differences among the models include their implica-

tions for currency and commercial loan behavior, the elasticity of money with respect to changes in base money, and the cyclic behavior of impact multipliers.

International Economics

KIYOSHI ABE, Ph.D. State University of New York (Binghamton) 1970. Dynamic models of economic development and international capital movements.

This study analyzes effects of capital imports on an underdeveloped economy aiming at rapid growth; effects of capital exports on an advanced economy desiring stable growth; and interacting effects of the capital movements on the respective economies in the context of an advanced-underdeveloped world. It concludes that the international capital movements can bring mutual benefits to both the advanced and the underdeveloped economy under specified plausible conditions.

FRANK J. ALESSIO, Ph.D. Oregon 1970. The international financial intermediation hypothesis: an empirical analysis.

PETER W. ANDREWS, Ph.D. Pennsylvania State 1969. The effect of import quotas on the United States zinc industry.

JACQUES ARTUS, Ph.D. California (Berkeley) 1969. A theoretical and empirical investigation of the short-run effects of domestic demand pressure on British export performance, 1954–67.

LLOYD C. ATKINSON, Ph.D. Michigan 1969. Toward a theory of dynamic comparative advantage.

V. N. BALASUBRAMANYAM, Ph.D. Illinois (Urbana) 1970. International trade in knowledge: a study of the Indo-foreign technical collaboration agreements in the private sector.

This study analyzes the role of international firms in transmitting technical know-how to Indian firms. Specifically the study focuses on the technical collaboration agreements that foreign firms have entered into since 1957. The growth in number of these agreements, the varied fields they relate to and their nature is specified.

RAVEENDRA BATRA, Ph.D. Southern Illinois 1969. Economic growth and the terms of trade.

The analysis of relation between economic expansion and terms of trade is approached by relaxing the specifications of the traditional model. New relations are established: Traditional results are reinforced if returns to scale are increasing, but rendered indeterminant if diminishing; Factor market distortions do not interfere with the direction of shift in the growing country's terms of trade subject to provisions. If factor growth is ultra-export-biased, the $3 \times 3 \times 3$ model leaves existing results unscathed. Positive elasticity of factor supply strengthens traditional analysis. Negative elasticity may reverse current results.

CARL F. BERGSTEN, Ph.D. Fletcher School 1969. The international roles of the dollar: focus for international monetary policy.

HUNTLEY H. BIGGS, Ph.D. Vanderbilt 1970. The effects of exchange devaluation on agricultural prices and costs in Argentina since 1959.

GEORGE K. CHRISTOU, Ph.D. North Carolina State 1970. Tariff structure and effective protection: the case of Greek manufacturing industries.

The study estimates and analyzes effective rates of protection for Greek manufacturing industries, and examines the effect of tariff structure on past allocation of resources in import substituting industries. Joint consideration of indirect taxes and tariffs is also made.

STEPHEN D. COHEN, Ph.D. American 1969. International monetary reform, 1964–68: a political interpretation.

The dissertation's central thesis is that the recently completed international monetary reform exercise which produced the Special Drawing Rights agreement is ultimately political in nature and a manifestation of the changing balance of power in Atlantic Community politics. After an initial theoretical analysis of the political overtones of international finance, the paper empirically examines the course of the monetary reform talks, specifically relating their development to the political dynamics of the Atlantic Alliance.

TERRENCE R. COLVIN, Ph.D. Massachusetts Institute of Technology 1969. A closed model of the U.S. balance of payments.

A closed model of the U.S. balance of payments is presented that includes capital and government accounts. Particular emphasis is placed on the dynamics of adjustment and on the simultaneous determination of the accounts. Attempts are made to ensure that the approach taken provides properties that various bodies of economic theory suggest are relevant. Policy implications of the model are pointed out during the course of the analysis. The model is used as a vehicle for presenting a new approach to the determination of the size and causes of various balance-of-payments deficit measures. This approach is compared to several currently used measures of the deficit.

BENOIT L. DAIGLE, Ph.D. Catholic 1969. The possibility of trade between Canada and Latin America.

The study of the imports of thirteen Latin American countries covering the period 1959–65 reveals that the region's demand for Canadian products has been limited to a few commodities. Canada's participation in the region market tends to be limited because of the following reasons: distance, high tariff and non-tariff barriers, political and economic stability, limited capacity of import, and little encouragement on the part of Canada for the region's commodities. Canada would benefit by promoting new foreign aid programs which

would improve the Latin American capacity to import and in drawing new bilateral trade agreements with major importers of the region.

WAYNE J. DAVIS, Ph.D. Rutgers 1970. The feasibility of preferential access to the United States for Latin American manufactured products.

This study examines the issues involved in the present controversy over preferences for the less developed countries. The dissertation traces United States' and Latin American attitudes toward preferences, analyzes the effectiveness of existing preferential systems, and evaluates the relative merits of preferences for Latin American manufactured exports to the United States.

JOHN T. DONNELLY, Ph.D. Vanderbilt 1970. External debt and economic development in postwar Brazil, 1947–66.

This study analyzes the causes and consequences of the highly unfavorable evolution of Brazil's external public debt in the postwar period. It is shown that the growth, composition, and terms of the external financing which facilitated the remarkably high GDP growth rates during the Kubitschek years were such that, by the early 1960's, a heavy burden of external debt service payments became a major factor initiating the downward adjustment in economic growth rates.

LAWRENCE P. DONNELLEY, Ph.D. Brown 1970. International capital movements, the terms of trade and Australian economic growth, 1861–1929.

A neoclassical one-sector growth model is the basis for analysis of international capital movements and Australia's economic growth. The model describes the time path of capital flows in terms of several parameters including the saving ratio, terms of trade, interest rate, rates of labor force growth and technological change. Evidence of the model's validity is provided by comparing predicted time paths with observed paths relative to predictions of a Harrod-Domar model.

KAMAL J. DOW, Ph.D. Missouri 1969. Latin American economic integration and the foreign trade of Colombia.

RICHARD D. DUCKWORTH, Ph.D. Colorado 1969. Special technical aspects of international emergency food management: their development during and after World War II and their present-day relevancy.

NAIEM A. EL-SHERBINY, Ph.D. California (Berkeley) 1969. Comparative advantage and development planning under the foreign exchange constraint, with special reference to Egypt.

WILFRED J. ETHIER, Ph.D. Rochester 1970. Economic growth, non-traded goods, and international trade and capital movements.

RICHARD A. FARRAR, Ph.D. New York 1969. Changes in concentration of Latin American trade, 1953–65.

MOHAMMAD A. FEKRAT, Ph.D. Indiana 1969. International monetary reserves for an expanding world economy: a theoretical inquiry.

This thesis suggests a new theoretical apparatus for determining the optimum international reserve requirements of monetary authorities. After a critical examination of two distinct approaches to the theory of demand for international reserves, a new line of analysis is developed along the Keynesian liquidity preference theory. Of the three motives outlined by Keynes, the precautionary motive is found to be dominant in the desire of monetary authorities to hold international reserves. Extending the implications of the foregoing analysis, it is shown that an inventory model could be developed to arrive at the optimum reserves to be held by monetary authorities.

JAMES W. FOLEY, Ph.D. Michigan State 1969. The balance of payments and import substituting industrialization in Argentina, 1945–61.

The study tests the hypothesis that balance of payments pressures induce import substitution. Techniques are developed to define operationally balance of payments pressure and to measure import substitution. Pressure years are identified as 1955–58. Extensive import substitution is seen to closely follow these years. The dollar cost of substituted products fell from almost half of total imports in 1956 to less than one-fifth in 1962, thus supporting the hypothesis for the Argentine case. Due to structural rigidities of the economy foreign investment was required to implement import substitution.

GLEN R. FOSTER, Ph.D. Texas (Austin) 1970. Automatic adjustment to international imbalance: an empirical investigation of the classical mechanism.

The classical process of adjustment to international disequilibrium, known as the price-specie-flow mechanism, was tested for validity (by using simple correlation techniques) with respect to the historical experience of three countries operating under different institutional arrangements: Mexico, a non-key-currency developing country; France, a non-key-currency developed country; and the United States, a key-currency developed country. The adjustment process tested consisted of four steps: balance-of-trade influences on reserve holdings (primary gold); reserve influences on the money supply; money-supply influence on domestic prices; and price influences on the merchandise trade balance.

CHARLES FREEDMAN, Ph.D. Massachusetts Institute of Technology 1969. Long-term capital flows between the United States and Canada.

This study develops a two-country model of international capital movements based on the portfolio behavior of American lenders and Canadian borrowers. The effects on interest rates and on capital movements of central bank operations and of the growth of

wealth in each country are derived. The model is then applied to the long-term capital flows between the United States and Canada for the years 1954 to 1966.

BERNARD FRIEDMAN, Ph.D. Columbia 1970. The impact of U.S. cotton policy on underdeveloped cotton-producing countries, 1946–65.

Using a model of U.S. raw cotton policy as a buffer-stock arrangement controlling both domestic and international prices, the study analyzes the effect of U.S. efforts to increase exports and dispose of excess stocks by government-financing programs and by direct subsidy of cotton exports. Policy and output responses of other producing countries are examined, and the responsibility of U.S. policy is appraised.

CHRISTOPHER GARBACZ, Ph.D. Iowa 1969. The problem of polarization in economic integration: Latin America.

JAMES D. GAVAN, Ph.D. California (Berkeley) 1969. Import substitution and exports: a case study of the pulp and paper industry in Chile.

ELMER R. GOODING, Ph.D. Kansas 1969. A general equilibrium analysis of the economic impact of discriminatory tariff reductions.

ROBERT GOODMAN, Ph.D. Michgian State 1970. An evaluation of the effect of the Dillon Round on the unit value and volume of U.S. imports and exports.

The objective of this investigation is to measure the effects of multilateral tariff reductions on the foreign trade sector of the American economy. In addition to yielding figures on price and quantity changes on a disaggregated basis, estimates are made of the terms of trade, welfare and employment effects of the negotiations. We conclude that the Dillon Round had a very limited effect upon the American economy.

JOHN K. GREEN, Ph.D. Virginia 1969. The international role of the pound sterling: its benefits and costs to the United Kingdom, 1957–67.

GEORGE B. HENRY, Ph.D. Yale 1969. Domestic demand pressure and short-run export fluctuations.

YUTAKA HORIBA, Ph. D. Purdue 1969. Interindustry analysis of the pattern of international and interregional trade.

A general model of trade which incorporates two economies characterized by open input-output-systems is proposed. As a special case, an input-output version of the Heckscher-Ohlin theory of trade is constructed under two alternative definitions of relative factor position. Non-linear programming techniques (Kuhn-Tucker conditions) are applied to deduce necessary conditions for full employment in terms of limiting values of the ratio of the marginal utilities and of the capital-labor endowment ratio. The empirical content of the theory is in turn examined in the context of Japanese interregional trades.

THOMAS O. HORST, Ph.D. Rochester 1970. The effect of tariff structures on the location of industry.

DAVID HUMPHREY, Ph. D. California (Berkeley) 1969. The effective rate of protection in Argentina.

LEODEGARIO M. ILAG, Ph.D. Purdue 1970. An econometric analysis of the impact of the U.S. sugar program on the Phillippine sugar industry.

This study attempts to measure quantitatively the impact of the sugar program by examining the effect on the area devoted to sugarcane yield, domestic consumption of centrifugal sugar, volume of sugar exports, and the induced level of technology. The analyses were made possible by a supply response and derived domestic demand models which were developed and analyzed with empirical data of 1950–60 for supply response and 1951–60 for derived domestic demand.

TAYSEER A. JABER, Ph.D. Southern California 1969. A theoretical analysis of dynamic aspects of economic integration of less developed countries.

MOHAMED A. JAMJOOM, Ph.D. Southern California 1970. International trade and balance of payments in a mono-product economy: a case study of the Saudi Arabian Kingdom.

In view of the analytical study of this dissertation, it can be concluded that: complete dependence on the petroleum industry is a very hazardous matter; plans should be made for diversifying the economy; real diversification can only be materialized by developing industries related to the abundant oil, natural gas, and minerals; external trade and foreign investments are essential for achieving this objective.

MARSHALL J. JEANNERO, Ph.D. Fletcher School 1969. The question of Mexico's economic dependence on the United States relative to the rest of the world, 1945–62.

JAMES E. JONISH, Ph.D. Michigan 1969. Collective wage determination and international trade: the U.S. and Canadian steel industry.

SEUNG HEE KIM, Ph.D. New York 1969. Korean balance of payments and economic development.

SHIN HAING KIM, Ph.D. Columbia 1970. Optimum tariff in the two-sector neoclassical growth model.

As an attempt to resolve the conflict between the classical trade and the modern growth theory in their implications for the most efficient allocation of resources, an optimum tariff model is constructed in the two-sector neoclassical growth model. The optimum rate of a tariff found in our model is necessary to satisfy the efficiency condition which has an analogous form with the Dorfman-Samuelson-Solow dynamic efficiency condition. Further, the optimum tariff path in our model satisfies the catenary property of the optimal path conjectured in The Turnpike Theorem and

fulfills the principle of The General Theory of Second Best.

JERRY L. KINGSTON, Ph.D. Pennsylvania State 1969. The instability of export proceeds of selected Latin American countries: 1948–65.

ROBERT C. KONWEA, Ph.D. Southern California 1969. European economic integration and African States.

This dissertation is an appraisal of the politicoeconomic consequences for African states of the evolution of the European Economic Community, European Free Trade Association, and Council for Mutual Economic Assistance. The EEF-African Association is essentially a customs union in which private investment and economic assistance are linked. Although the Community brings real benefits to both its European and African members, it reinforces old colonial demarcations and retards African unity.

KENNETH J. KOPECKY, Ph.D. Brown 1970. Growth and developed economies: a study of the OECD countries.

The thesis examines long-run international capital movements in the context of a single-sector open economy growth model with a fixed world interest rate. The model is solved for the ratio of net factor income (repayment on past borrowing) to gross national product in terms of parameters such as the savings ratio, growth rate, and capital coefficient. Application of the model to the OECD countries indicates acceptance of the growth model hypothesis.

ELIAHU S. KREIS, Ph.D. Oklahoma 1970. Voluntary controls on direct investment abroad and the U.S. balance of payments, 1962–67.

This study describes the balance-of-payments problem confronting the United States during the years 1962–67; the short- and long-term balance of payments adjustment approaches are the Voluntary Restraint Program or direct investment abroad during the years 1965–67. Primary emphasis is placed on the analysis of the effects on the balance of payments of restraints on direct investment abroad. Attention is given to the factors which are directly related to capital outflows from the United States and increased investment abroad. In particular, the study measures the direct investment recoupment periods as related to the balance of payments. Finally, the study assesses the success of the program in relation to the stated objectives.

PETER K. KRESL, Ph.D. Texas (Austin) 1970. The responsibilities of surplus countries in balance of payments adjustments: West Germany, a case study.

We have developed a policy matrix, incorporating relative inflation as its major argument, which is appropriate, given current institutional arrangements, to surplus countries. This matrix was then used to formulate a critical evaluation of recent economic policy in West Germany, the world's leading surplus country. We find that by ignoring the dictates of the Gold-Exchange Standard system West Germany has contributed to the crisis-environment which has characterized international monetary relations for the past decade or so.

ANTHONY M. J. LANYI, Ph.D. California (Berkeley) 1969. Economic welfare under fixed and flexible exchange rates: a study in the theory of economic policy.

KUO-SHU LIANG, Ph.D. Vanderbilt 1970. Foreign trade and economic development in Taiwan, 1952–67.

DONALD L. LOSMAN, Ph.D. Florida 1969. International economic sanctions: the boycotts of Cuba, Israel, and Rhodesia.

This thesis studies the economic impact of international sanctions upon three target nations—Cuba, Israel, and Rhodesia—testing the proposition that economic sanctions can cause sufficient disruptions of a total economy to bring about the political changes desired by the boycott initiators. The thesis found little support for the hypothesis tested. Although the nations studied all suffered economic damage, sometimes quite considerable, the potential for altering trade patterns and for domestic adaptations are crucial variables either overlooked or underestimated by sanctions initiators.

JOSEPH A. MCKINNEY, Ph.D. Michigan State 1969. The economic nature of East-West trade.

This study deals with the effects of Soviet-type economic planning upon the commodity composition of East-West trade. The organization and planning of foreign trade is discussed, as well as domestic price formation. A Leontief-type of Soviet bloc trade with the West is conducted to see if this trade conforms to the postulate of the Hechsher-Ohlin model. Calculations are repeated to incorporate human capital into the analysis.

KATHERINE W. MIDDLETON, Ph.D. Fletcher School 1969. Eurodollars and U.S. commercial banks: the implications of the 1966–69 experience for domestic and international monetary policy.

TAPAN MUKERJEE, Ph.D. Colorado 1970. Economic impact of decolonization: the British-India case.

CHARLES CHAU-FEI OU, Ph.D. North Carolina (Chapel Hill) 1969. German banks' demand for short-term foreign assets and the monetary dependence of the German economy, 1960–67.

This is a statistical investigation of the loss of monetary independence of the German economy as a result of increasing connection between the German and the foreign money markets. Three main topics discussed are the statistical estimation of: the German banks' holdings of short-term foreign assets, the supply of money function, and the determinants of the

stock of money and short-term interest rates. The results indicate that the Bundesbank's policy has been ineffective in affecting the money supply and the short-term interest rates.

KENNETH M. PARZYCH, Ph.D. Connecticut 1970. An analysis of export trade activity under the Webb-Pomerene Act of 1918.

MONIQUE PAUL-GARRITY, Ph.D. Boston College 1969. The pattern of coffee imports in the common market countries with special reference to the association between the EEC and the African States.

Contrary to predictions, a statistical analysis reveals that no diversion took place in favor of the African Associated States between 1958 and 1966 as a result of the preferential margin accorded to these states. Relative prices played a minor role in explaining the relative quantities of coffee imports from competing sources and that other factors such as consumer preferences and buyer-seller ties were much more important.

DONALD K. PEMBERTON, Ph.D. Kansas 1969. An empirical test of a model of the foreign exchange market.

HUGH M. PINCHIN, Ph.D. Yale 1970. Canadian tariff levels 1870–1959.

USMAN A. QURESHI, Ph.D. Houston 1969. Some empirical tests of the classical comparative cost theory.

The classical comparative cost theorem is empirically tested by utilizing data for the United States and the United Kingdom, Australia and Canada, and Argentina and the United States. The statistical estimation procedure employed is simple and multiple regression of data. The evidence presented supports results of the previous empirical studies of MacDougall, Stern, and Balassa for countries at similar levels of economic development. However, inconclusive results are obtained for countries at dissimilar levels of development.

MOHAMMAD RAQUIBUZZAMAN, Ph.D. Cornell 1970. An economic appraisal of the sugar policies of developed countries and the implications of these policies to developing nations.

This study outlines alternative sugar programs which could be adopted by developed countries, and examines the economic implications of these policies to major sugar exporting nations. Estimates of world supply, demand, and price of sugar through 1980 were made under the assumption of completely free trade. Given such a policy, total gains to developing countries were estimated at $1,398 million. If industrial countries adopted a policy which limited their production to the 1964–66 level, and imported subsequent demand increases, the resultant gains to developing countries would be about one billion dollars by 1980.

EDWARD J. REGAN, Ph.D. Fordham 1970. Domestic stabilization policies under alternate exchange rate assumptions: a comparative analysis.

STEPHEN E. REYNOLDS, Ph.D. Wisconsin (Madison) 1970. Concentration of trade and the instability and growth of exports: developing Asia.

The purpose of this study is to assess the impact of concentration of trade on the instability and growth of countries' exports. Concentration of exports and of imports is measured. Export instability is defined and attention is given to problems of measuring it. The study uses linear regression techniques and cross-sectional data for the ECAFE countries (and for some developed nations). The implications of the statistical results for export diversification policies are discussed.

RITA M. RODRIGUEZ-MEDEROS, Ph.D. New York 1970. Fluctuations in U.S. imports 1948–66: Economic analysis and econometric model.

HENRY S. ROSENBERG, Ph.D. Arizona 1969. Capital inflow and the economic development of modern Israel.

THOMAS A. SEARS, Ph.D. Harvard 1970. United States exports and domestic expenditure: the short-run relationship.

A rapid upswing in domestic expenditure put pressure on domestic resources, causing the supply of exports to be restricted. The effects on U.S. merchandise exports were explored by means of multiple regression analysis on quarterly data, 1958–68. Lengthening delivery times for durable goods and increasing rates of capacity utilization in manufacturing reduced exports. In addition to these restraining effects, foreign trade multiplier feedbacks were also considered. These positive feedbacks only partially offset the restraining effects in the first respending round, but ultimately tended to dominate them.

ABUL K. SIDDIQUE, Ph.D. Yale 1970. The international monetary and economic conference of the inter-war period: a study in international cooperation.

GHAZI A. SIRHAN, Ph.D. North Carolina State 1969. An elasticity of substitution and a market share approach to the British and German import demand for U.S. cotton.

Two models, an elasticity of substitution and a market share approach, were used to investigate two foreign markets for cotton. From these estimates some implications for direct price elasticities can be derived. The relation between cotton and synthetics is also explored.

HOUSTON H. STOKES, Ph.D. Chicago 1969. The adjustment mechanism in the interest arbitrage market: a theoretical and empirical study.

This dissertation investigates the adjustment process in the foreign exchange market within the context of the interest parity theory. Tsiang's proof was shown

to be invalid under some conditions of government intervention. Adverse changes in the balance of trade were shown to worsen the arbitrage outflow in some fixed exchange rate situations. An index of tension (or crisis) in the foreign exchange market was developed and tested in the period 1921–69.

JUSTIN D. STOLEN, Ph.D. Illinois (Urbana) 1969. A study of certain motivations behind long-term portfolio capital flows between countries: 1958–67.

The purpose of this study is to obtain more specific estimates of determinants of international portfolio capital flows between the United States and certain European countries. Specifically the study looks at both price and non-price motivations; integrates them on a country by country basis, and determines their relative importance.

PETER S. STOWE, Ph.D. Purdue 1970. The foreign exchange market: three models.

This dissertation attempts to contrast the movement of short-term capital and the central bank policies of pegging the spot rate, intervening in the forward market, and changing the interest rate, in three two-country models. The first model has exogenously determined interest rates, the second is a Euro-dollar model, and the last has endogenously determined interest rates.

LANNY E. STREETER, Ph.D. Illinois (Urbana) 1970. Optimal international reserve holdings: an inventory model and empirical tests.

The study develops a stochastic model for determining an optimal level of international reserves for a country when reserves serve the sole purpose of financing temporary external payments imbalances for various countries. In addition, I estimate a cross-sectional reserve demand function based on the theoretical model.

FRANK STUBENITSKY, Ph.D. California (Berkeley) 1969. American direct investment in the Netherlands industry: a survey of the year 1966.

NABIL SUKKAR, Ph.D. Indiana 1969. Chenery-Bruno test of Egypt's first comprehensive plan.

Chenery's gaps model is applied to the Egyptian economy. The model is used to evaluate the first five-year plan (1960/61–1964/65) and to make projections for a future plan.

CLYDE P. TOPPING, Ph.D. New York 1970. The economic desirability of a Lafta customs union.

WILLIAM G. TYLER III, Ph.D. Fletcher School 1969. Balance-of-payments adjustment in less developed countries and the International Monetary Fund.

WILLIAM P. WADBROOK, Ph.D. Fletcher School 1970. Formulation and administration of German balance-of-payments policy, 1961–69.

GEORGE C. WANG, Ph.D. Columbia 1970. A quantitative analysis of income and scale effects on imports: 1961–65.

This is a multiple regression analysis of the effects of per capita income and scale on total imports. Data were collected from 72 countries for the period 1961–65. To test empirically the Heckscher-Ohlin theory, countries are grouped as labor-, capital-, or land-abundant; while imports are classified as labor-, capital-, or land-intensive, then expected imports are compared with observed imports. The results seem to suggest that the theory is not untenable.

DENNIS J. WEIDENAAR, Ph.D. Purdue 1969. The impact of the Central American common market on Guatemala's trade flows.

This study is an attempt to determine the impact of the Central American common market during the period 1959 to 1965 on the regional trade flows of Guatemala. The conclusions are that regional tariff liberalization has served to: restructure regional imports in favor of high pre-union tariff commodities, restructure extra-regional imports in favor of low pre-union tariff commodities, and generate no apparent effect on restructuring high pre-union tariff regional imports in favor of regional imports.

GOPAL JI YADEV, Ph.D. Queen's 1970. The discriminatory aspects of Canada's imports of manufactured goods from the less developed and the developed countries.

This study provides a framework to measure discrimination in Canada's import policy. This study is limited to the investigation of discrimination against the imports of manufactures in 1964. There are two main conclusions: The evidence suggests that Canadian import control methods discriminate against the imports of manufactured goods from the less developed countries. Canadian import control methods discriminate more heavily against the imports of finished manufactures from the less developed countries, than against the imports of semi-finished manufactures.

LAWRENCE ZIEGLER, Ph.D. Iowa 1969. Monetary accommodation of regional integration: Latin America.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting

DAVID A. AAKER, Ph.D. Stanford 1969. The long-term value of temporary price reductions.

HAROLD C. ALLEN, Ph.D. Florida 1969. Determining the information system required for effective management of the redesign of aircraft engine parts.

System research with respect to automation (SAM) is suggested as an operational theory approach toward improved problem definition and solution of a press-

ing air transport problem of excessive time required in the present design-redesign cycle of aircraft engine parts. Safety and airline profits are affected by excessive cycle time of after-product support redesign.

TOM C. ALLEN, Ph.D. Florida 1970. A theoretical approach to risk management.

A model was constructed to simulate the risk management decision process of determining when a bundle of pure risk should be shifted or retained. The cost and benefit aspects connected with the various pure risk treatment techniques was determined by an analysis of such factors as the cost of excess liquidity, expected losses, administrative and adjustment expense, premium savings, and the risk aversion attitude of management.

CLAYTON W. ANDERSON, Ph.D. Northwestern 1969. An examination of the historical and theoretical foundations of local government accounting.

This study explores the increasing importance of managerial accountability in local government administration. It highlights four limitations of local government budgeting, two serious shortcomings of accounting practices, and three basic shortcomings in accounting principles. The study recommends that supplementary full-cost statements of cost of operations be prepared for each general purpose fund and each organizational sub-entity within each fund. Further, a consolidated statement of cost of operations should be prepared for the general and special revenue funds.

EVAN E. ANDERSON, Ph.D. Cornell 1970. Measurement of market and brand effects.

WILTON T. ANDERSON, JR., Ph.D. Michigan State 1969. An analysis of the correlates of convenience-oriented consumer behavior: with special emphasis on selected convenience foods and durable goods.

The research objective was to distinguish consumer typologies emerging from patterns of convenience food consumption and convenience goods ownership or usage through the application of chi-square and hierarchical cluster analysis. Findings derived from mail questionnaire responses and comprehensive demographic data for 796 families from a representative panel of 1,000 suburban families revealed distinct typologies, differentiated by stage in the family-life cycle and selected indicators of socioeconomic status, yielding significant implications for market segmentation.

VINCENT P. APILADO, Ph.D. Michigan 1970. Industrial aid bonds: an inquiry into their use and significance.

ALLAN BAILEY, Ph.D. California (Los Angeles) 1969. A computerized standard cost and variance analysis model for improved cost control.

VIRGINIA H. BAKAY, Ph.D. Alabama 1969. An analysis of legal claims against certified public accountants for liability related to the audit function.

L. EUGENE BALDWIN, Ph.D. Florida 1969. Differentiation, integration, and performance in selected Florida hospitals.

An empirical study of 14 similar Florida hospitals is made to investigate the relationships of differentiation characteristics and integrative activity in hospital organizations to organizational performance. A descriptive analysis is used to construct an index of evaluating overall hospital performance. The ranking of the sample hospitals by performance rating is compared to sample ranking by differentiation deviations and by quality of integrative activity. The findings suggest a significant relationship between organizational differentiation and integrative characteristics and organization performance.

GUY R. BANVILLE, Ph.D. Alabama 1969. An analysis and evaluation of source selection criteria of private residential builders in standard metropolitan statistical areas in the Southeast.

PAUL R. BEACH, Ph.D. Brown 1970. Equity values and corporate debt in the capital structure.

A model is developed in which it is assumed, among other things, that the firm's real investment and earnings grow over time. The use of corporate debt in each period over an infinite horizon leads to both an increase in the expected stream of dividends and risk. To determine which factor dominates, regression analysis is applied to data on selected firms. In some cases, equity values increase with addition of debt in the capital structure.

RICHARD E. BELL, Ph.D. Carnegie-Mellon 1970. A theoretical investigation of optimal firm financial behavior under conditions of uncertainty and imperfect capital markets.

CARL J. BELLAS, D.B.A. Oregon 1969. Corporate performance and managerial attitudes in worker-owned plywood mills.

JACOB BEN-MOSHE, Ph.D. New School 1968. Factors affecting the growth of the pharmaceutical industry in India.

CHARLES O. BETTINGER III, Ph.D. Texas (Austin) 1969. Estimation and prediction of interzonal traffic flows using quantitative models.

Extremal boundary methods can be applied to problems of human interaction where the individuals involved are separated by distance. A history of classical solutions and criticisms is presented as well as the predominant methods now found in the literature. Proof is given which shows that the gravity model, the most widely used mathematical formulation, is equivalent to a particular non-linear optimization model. Mathematical proofs are shown as well as a

brief example comparing the results of the two models.

WAYNE D. BODENSTEINER, Ph.D. Texas (Austin) 1970. Information channel utilization under varying research and development project conditions: an aspect of inter-organizational communication channel usage.

The purpose of the study was to enlarge the understanding of the interface between research and development organizations and the process of information transfer and diffusion within the scientific and technical communication system. The research focused upon the utilization of interpersonal communication channels between the transacting organizations of R&D projects under varying project conditions. The major premise was that the utilization of interpersonal communication between the transacting organizations of a R&D project is a function of the varying project uncertainties. This premise was partially confirmed.

ROBERT J. BOEWADT, Ph.D. Michigan State 1970. An analysis of the internal sales force as a factor in the design of total sales strategy.

This study focused on the inside salesman within the steel industry. The theoretical framework of role theory was used in defining the relative influence exerted by customers, field salesmen and managers on the inside salesman. Role consensus between the various role definers and the inside salesman was analyzed with respect to inside selling effectiveness. The salesman's internal role was similarly analyzed with respect to inside selling effectiveness. Environmental factors such as firm size, market and breadth of line were considered.

JOHN M. BURNHAM, Ph.D. Texas (Austin) 1970. Conditional chance-constrained programming in portfolio management.

The thesis is an application of mathematical programming techniques to the portfolio selection and adjustment problem. The objective is satisficing to maximum probability, rather than maximizing profit. Uncertainty is explicitly recognized and the probabilistic nature of both the future securities prices and the assets available for investment is included in the formulations.

MARY E. BUTCHER, Ph.D. Catholic 1969. The internationalization of the American hotel industry.

ARTHUR G. BUTLER, JR., Ph.D. Florida 1969. Behavioral implications for professional employees of structural conflict associated with project management in functional organizations.

JOSEPH CALTAGIRONE, Ph.D. New York 1970. The rate of return on equity capital: an econometric study.

ERNEST E. CARTER, Ph.D. Carnegie-Mellon 1970. A behavioral theory approach to firm investment and acquisition decisions.

ARTHUR B. COCANOUGHER, Ph.D. Texas (Austin) 1970. An analysis of nonmembership, socially-distant reference group influence on consumer aspirations.

Behavioral science theory has long held that a social group can influence the attitudes and behavior of nonmembers, even if there is no face-to-face interaction between the group and the nonmembers. The research project was undertaken to analyze the influence that one such potential reference group, the business executive group, has on the consumer aspirations of college students. The principal research hypotheses deal with the notion that the more attraction the nonmember feels toward the group, the greater will be the influence of the group over the consumer products desired by the nonmembers.

CHARLES J. COEN, Ph.D. Michigan State 1970. A field study of leadership in a complex organization.

This field study was designed to investigate and expand upon the Fiedler Contingency Model of leadership effectiveness, and it used three different leadership measures. The research subjects were a major university football team which has been successful during the fifteen-year tenure of the current head coach. The organization's results did not support the model's predictions for either the head coach or the total leadership group, however, methodological difficulties may have contributed to these findings.

BOYD D. COLLIER, Ph.D. Texas (Austin) 1970. The foundations of socioeconomic accounting.

This study shows that the discipline of accountancy should be more involved in our social and economic accounts. A national accounting system is developed which would improve the meaningfulness and increase the uses of national income data if it were implemented. This model of socioeconomic accounts is developed while making use of John Dewey's theory of valuation. This model also introduces the concept and measurement of disproduct into the national accounting system.

ROBERT B. CONRAD, Ph.D. North Carolina (Chapel Hill) 1969. Intermittent interference with pseudorepetitive operations and its effect on standard time and the learning curve.

The research attempts to show that the existence of learning and interdependence of work elements invalidates several assumptions which underlie the traditional time measurement process. A simulation measurement model was proposed to recognize the interdependence of work elements in the prediction of tasks which varied in sequence. Learning was also quantified and incorporated in the proposed measurement model to permit the model to predict over a wider range of performance.

NORMAN G. COURNOYER, Ph.D. Massachusetts 1970. An econometric model of the occupancy for hotel-motel rooms in the Pioneer Valley, Massachusetts.

WILLIAM K. CUNNINGHAM, Ph.D. Texas (Austin) 1970. A comparison of the relative effect of favorable and unfavorable information on judgmental choice conviction.

Studies in persuasive communications have resulted in reports of a preponderance of attitude change following receipt of persuasive messages. Large movements have been attributed, in part, to the lack of ego-involvement of the respondent. An experiment was devised which tested various aspects of relative influence of favorable and unfavorable, direct and indirect, information on the conviction of a nonego-involving choice, using an eleven-interval form of semantic differential scales. Results generally supported the contention that the greatest influence on conviction of such choices are the most direct and least subtle messages, regardless of whether they are favorable or unfavorable.

DUDLEY W. CURRY, Ph.D. Stanford 1969. The financial reporting of convertible debentures.

NEIL G. DAVEY, Ph.D. Michigan State 1969. Influences on an external consultant's effectiveness in assisting organizational change.

This research investigated the nature of the relationship between an organization and an external consultant it had retained with respect to the effectiveness of the ensuring consulting assignment. Empirical data were obtained from 133 organizational respondents which suggested several conditions which an organization considering and then proceeding with a consulting assignment should observe. Further evidence was developed which showed a relationship between the frequency of, and arrangements for, consultant assistance by an organization and a measure of the open-mindedness of the organization's chief executive.

HAROLD D. DEWHIRST, Ph.D. Texas (Austin) 1970. The socialization of the young professional: a study of changes in the career related values of engineers and scientists during the first five years of employment.

A study of professional and organizational values of engineers and scientists during their first five years of employment. More experienced employees are hypothesized to act as socialization agents in influencing the values of the young professionals. Values as a function of length of service are examined to determine if the values of the young professional become more similar to those of the more experienced employees. Relationship of interaction and the degree to which values are shared is examined.

PARKS B. DIMSDALE, JR., Ph.D. Florida 1970. A history of the cotton producers association.

The thesis is an analysis of the inception, growth, and development of one of the five largest farmers' marketing and purchasing cooperatives, during the period 1933 through 1968. While attention is given to the environment for cooperation, the examination centers on the organization, administration, and operation of the enterprise as it faced numerous challenges in seeking to meet the production and marketing needs of the farmer in the Southeast.

JAMES S. DYER, Ph.D. Texas (Austin) 1969. Cost-effectiveness analysis for a public system of higher education.

The purpose of this dissertation is to develop an approach for the use of cost-effectiveness analysis as an aid in planning for the higher education system of a state. As an illustrative example, two alternatives for the expansion of a higher education system are compared by estimating their associated benefits and costs. Although data and the stated objectives of the public system of higher education of the state of Texas are used, the proposed methodology is general in nature and should be adaptable to the needs of other states.

GEORGE H. EBBS, Ph.D. Columbia 1969. Sensitivity analysis as a tool for affecting the rational evaluation of corporate decision alternatives: a study of capital equipment planning processes in two U.S. airlines.

FRANK N. EDENS, Ph.D. Texas (Austin) 1969. A study of relationships of personality needs and biographical data to work activity choice in a selected group of physicists.

The purpose of this study was to investigate the relationships of personality needs and biographical data to professionals' choice of or preference for work activities. Two hundred and seventy-six physicists completed a psychological test and a questionnaire which included biographical items. Relationships were found between personality needs and both choice of and preference for work activities. Except for the rating of the graduate faculty in physics of the school from which the physicist received his highest degree, little relationship was found between biographical data and work-activity choice.

JOHN S. ELLETT II, Ph.D. North Carolina (Chapel Hill) 1969. Interstate corporate income taxation of truckers and movers.

The research analyzed the problems of interstate corporate income taxation of truckers and movers within the transportation industry, a service industry, from a legal and a taxpayer's compliance viewpoint. The conclusion suggested guidelines for any future legislation by Congress as well as recommendations for simplification of the compliance procedures for the corporate taxpayer operating in interstate commerce.

ROBERT C. ELLIS, D.B.A. George Washington 1969. Decision trees applied to capital investment and technological uncertainty: a case study.

The research question is whether the decision tree technique might have provided a better method for making a decision involving technological uncertainty than a decision process which primarily utilized man-

agerial judgement. Limited conclusions indicate a yes answer to the research question. Problems of decision tree application are analyzed and resolved, notably the difficulty of obtaining probability estimates from management. A sensitivity analysis alleviates, but does not remove, this problem.

ALI A. EL.-MEZAWIE, Ph.D. Texas (Austin) 1970. A mathematical model for the ultimate tow size for navigation of sharp bends on the lower Mississippi river.

The study brings into focus a problem regarding the restrictions of sharp bends on tow size and route capacity on the lower Mississippi River. A sharp bend creates two restrictions. First, it limits the length and breadth and hence the capacity of a single tow. Second, it creates a bottleneck situation when two large tows cannot pass. Consequently, traffic is limited to that which can move through the bend in a given time. The main purpose of this study is to find out the maximum time. The solution of the problem has been tentatively reached through a construction of a mathematical model derived from a fundamental law of physics ($F = Ma = M_R^{V^2}$).

EDWIN J. ELTON, Ph.D. Carnegie-Mellon 1970. Studies in financial theory.

PAUL E. ERZEN, Ph.D. Michigan State 1969. An empirical analysis of an Air Force item manager role and its relationships with automated processes.

This is a descriptive study of the Air Force economic order quantity item manager role and its relationships with the automated systems which provide information for performance of various stock control functions. Some basic role theory concepts were used to examine the item manager's role behavior and attitudes relating to the automated systems environment. Results indicated that a majority of item managers did not directly perceive automated systems as primary sources of role conflict.

THOMAS G. EVANS, Ph.D. Michigan State 1969. An approach toward the effective communication of internal accounting information for control.

A comprehensive set of control reporting guidelines is developed and recommended for achieving effective communication through control reports. It will eliminate and overcome the major barriers to effective communication which exist in the control reporting communication situations, such as too many reports being prepared and sent, a lack of feedback from the receivers, and receiver dissatisfaction with control report format.

ALFRED M. FALTHZIK, Ph.D. Michigan State 1969. Analysis of selected socioeconomic variables and their effect on consumer temporal behavior in different size shopping areas.

Major findings of the study indicate that the average amount of time consumers devote to shopping on a single shopping trip increases as the size of the shopping area increases. Significant differences exist between households of various income levels and different occupational categories.

EDGAR R. FIEDLER, Ph.D. New York 1970. Measures of credit risk and experience.

LYNDELL W. FITZGERALD, Ph.D. Purdue 1969. A theory of control for disassembly production processes: an application to the meatpacking industry.

The Production Control System consists of the following: A Production Demand Generator which develops weekly sales forecasts that are mapped onto daily production schedule; An Inventory Management Simulator which utilizes a modified linear programming algorithm to simulate order filling, detect stockouts, maintain an inventory account and control inventories; A Production Control Decision Analyzer which uses an LP scheduling algorithm and other subsystems input to develop detailed production plans.

MARTIN S. FREDERIXON, Ph.D. Michigan State 1969. An investigation of the product life cycle concept and its application to new product proposal evaluation within the chemical industry.

In an effort to integrate product life cycle and capital budgeting theories, structural characteristics of new industrial chemical products were related to performance with an emphasis on classifying product life cycle patterns. It was also possible to use historical experience as a basis for making a number of recommendations on the handling of the commercial development process.

PAUL M. FRISHKOFF, Ph.D. Stanford 1970. An empirical investigation of the concept of materiality in accounting and its application in financial reporting.

EARL FOSTER, Ph.D. New York 1969. Analysis of common stock prices.

STEPHEN H. GAMBLE, D.B.A. Oregon 1970. The efficiency and effectiveness of the despensa system of food distribution in Monterrey, Mexico.

LOUIS GANZ, Ph.D. New York 1969. An empirical examination of the relation between capital structure and business risk.

DAVID M. GEORGOFF, Ph.D. Michigan State 1969. The effects of odd-even retail pricing on value determination, product perception, and buying propensities.

The study is a two-stage effort. The first is a simulated inquiry which measures the effects arising from odd-even retail prices on the value estimates, perceptions, and purchase dispositions for a group of eleven department store items. The second uses an experimental design to test sales for the same products at alternative odd-even endings for a four-week period in six stores of a department store group.

PETER L. GILLETT, Ph.D. Michigan State 1969. An analysis of demographic, socioeconomic and attitudinal characteristics of the urban in-home shopper.

The research analyzed the relationship among selected characteristics of urban female shoppers and their telephone and mail shopping for general merchandise. The following problem areas were investigated: Influence of perceived and actual "locked-in" shopping conditions on decisions to purchase by phone or mail; socioeconomic and demographic characteristics discriminating in-home shoppers from other shoppers; extent and influence of "convenience orientation" on decisions to shop in-home; relationships of attitudes toward shopping to decisions to purchase at home.

EDWARD GOLDBERG, Ph.D. Columbia 1969. Operations research: a new specialty within business organizations.

NICHOLAS J. GONEDES, Ph.D. Texas (Austin) 1969. Accounting for common stockholders: a decision making and motivational approach.

The purpose of this dissertation is to develop and explain a behavioral model that is relevant to accounting. The behavior of common stockholders constitutes the axial phenomenon of the propositions set forth. The content of this dissertation embraces: a general model of behavior that is derived, mainly, from the theoretical constructs of *field* and *gestalt* psychology; the psychological aspects of problem solving, i.e., decision making; and the psychological and philosophical aspects of expectations, uncertainty, and behavioral changes.

EDDIE E. GORDHAMMER, D.B.A. Oregon 1970. A discriminant analysis mode for rating research and development programs.

This research developed and demonstrated a discriminant analysis method for rating technical data associated with Department of Defense (DOD) and National Aeronautics and Space Administration (NASA) research and development programs. It was designed to identify any differences in the technical data rating philosophies of the participating decision-making groups. Questionnaires were sent to middle-level engineers and other technical personnel within the research and development industrial-government aero-space complex.

WILLIAM P. GREEN, Ph.D. North Carolina (Chapel Hill) 1969. A study of commercial bank size and small business financing.

DANIEL GREENO, Ph.D. California (Los Angeles) 1969. The nature of signal value and the orienting reflex.

THEODORE F. GROVES, Ph.D. California (Berkeley) 1970. The allocation of resources under uncertainty: the informational and incentive roles of prices and demands in a team.

CHARLES M. GUDGER, Ph.D. Kansas 1970. The effects of sex differences upon opinion change.

A study of the effect of three independent variables in the communication process (communicator, message, and recipient) on the dependent variable—effectiveness (as measured by opinion change).

THOMAS L. GUTHRIE, Ph.D. Purdue 1970. Initial forward planning strategy and successive review by the commercial feed firm utilizing a static compacted linear programming model.

This study was designed to construct and test a computerized operational control system (OCS) for commercial feed firms. The OCS simultaneously evaluates procurement, production, and sales activities in generating optimum decision guides which may be used for overall planning and control by top management. A matrix compaction technique permitted an 80-90 percent reduction in the number of constraint equations over traditional models.

HERBERT R. HAHN, Ph.D. North Carolina (Chapel Hill) 1969. The effect of interest rate restrictions on the flow of mortgage funds.

KENNETH G. HARDY, Ph.D. Michigan 1969. An experiment to study the taste experience on ability to discriminate and preference reliability in food acceptance testing.

DOROTHY N. HARLOW, Ph.D. Kansas 1970. The promotional preference of professionals: an empirical test of eclectic theories.

The purpose of this study was to gain a better understanding of why some highly skilled, well-educated people, often referred to as professionals, choose to go into management while others prefer to continue working in their areas of specialization.

WILLIAM S. HART, Ph.D. Florida 1969. Early and current management theorists: a comparative functional analysis.

This study compares the principles of selected management theorists to determine the extent to which common principles have been developed by the early theorists. The method used was to construct an analytical framework of basic ideas within the general areas of the management functions to categorize and contrast their respective ideas.

NABIL M. A. M. HASSAN, Ph.D. Alabama 1969. Significant contributions of George Hillis Nerlove to the development of accounting thought.

JOSEPH P. HAWRANEK, Ph.D. Pennsylvania 1970. An operational means of detecting potential sources of theft in a retail-warehousing environment.

The purpose is to demonstrate a technique that allows more meaningful exception reports to be supplied in a previously "unmeasurable" area. The scope is limited to the detection of potential sources of cullusive theft in remote sales units. The mathematical

techniques used are Bayesian inductive inference, Shannon's mutual information measure, Theil's information theoretic inequality measure and non-parametric median detector measures, computer simulations were used to test the validity of the derived models.

RONALD L. HEIM, Ph.D. Cornell 1969. An integral formulation of purchase allowances into deterministic and stochastic inventory systems.

The study develops an inventory control system for supermarket purchasing. The uniqueness over traditional systems lies in the treatment of nonrecurring purchase allowances. A system of response measurement to promotions is developed along with an estimated brand loyalty index. The derivation of optional purchase quantity equations and the quantification of purchase allowances achieves objectivity in supermarket purchasing decisions. The technique was developed using a two-year historical base of the weekly movement of 328 grocery items.

MURRAY S. HENDERSON, Ph.D. California (Los Angeles) 1969. Some factors influencing the annual reports of North American corporations.

THOMAS E. HENDRICK, D.B.A. Oregon 1970. A heuristic model for client oriented agencies: a parametric linear programming design with a priority-utility objective function.

This study formulates and tests a program budgeting model for a client-oriented public agency through a parametric linear programming design which utilizes a priority-utility objective function. This research involved development of an experimental methodology based upon the Churchman-Ackoff Successive Comparisons method to capture a relative priority-utility function among nine client programs according to eight carefully defined criteria which reflected agency objectives. The priority-utility function which resulted from these experiments forms the objective function in the linear programming model.

DAVID F. HITCHCOCK, Ph.D. California (Los Angeles) 1969. A comparison of methods for determining the true price of life insurance.

THOMAS R. HOFSTEDT, Ph.D. Stanford 1970. Some behavioral implications of aggregation in accounting reports.

A. THOMAS HOLLINGSWORTH, Ph.D. Michigan State 1969. Relationships between the informal organization and the effectiveness of formal leaders.

This study was designed to examine the relationships between the formal leader's perceptions of the informal organization within his work group and his effectiveness as a formal leader. The study demonstrated a significant association between an accurate perception of the informal organization and a high degree of formal effectiveness. There was no significant association demonstrated between the formal leader's perceived degree of control over the informal organization and his formal effectiveness. There was also no significant association demonstrated between a leader's membership in the informal organization and his effectiveness.

JOHN W. HOLSINGER, Ph.D. North Carolina (Chapel Hill) 1969. EDP implementation: an analysis of constraints.

JOAO C. HOPP, Ph.D. Michigan State 1969. A study of the investment policy and performance of subsidiaries of U.S. manufacturing corporations in Brazil.

The purpose of this study is to examine and evaluate the investment performance and policies of subsidiary companies in comparison with their parent corporations. The basis hypotheses are: The subsidiary companies are net monetary debtors; investment in Brazil carried more risk than in United States; a better performance, as measured by profitability ratios, will be obtained by subsidiaries in comparison with their parent companies. It was found that subsidiary companies are not net monetary debtors, Brazilian investment does carry more risk than investment in United States, and subsidiary companies do not have better performance.

CHARLES W. HUBBARD, Ph.D. Arkansas 1970. An examination of the variations in sales performances of manufacturer and distribution records of supermarket items under different merchandise shelf arrangements.

BILLY R. HUMPHREY, Ph.D. Arkansas 1970. An evaluation of selected decision techniques in relation to the post-audit of capital expenditures.

EDET R. IWOK, Ph.D. Northwestern 1970. An application of economic service potential method of depreciation in accounting: a theoretical model.

The decline in the service potential of a fixed capital asset is a rational basis for depreciation accounting. The service potential of the asset is measured by the use of a theoretical second-hand asset value. This theoretical value is defined as the price which would make the owner of an old asset indifferent between selling his asset and holding it. The depreciation for any period is the decline in the service potential of the asset determined by the difference between the theoretical second-hand asset value at the beginning and at the end of the period.

TRIBHOWAN N. JAIN, Ph.D. Michigan State 1970. A study in the effects of alternative methods of accounting for income taxes on term loan decisions.

This dissertation studies the effects of alternative methods of accounting for income taxes used in the preparation of financial statements on term loan decisions by commercial banks. The results of the study indicate that the differences in accounting methods influence the decisions of the loan officers. These officers use surrogated data and they usually rely on the auditor's opinion for the applicability of generally ac-

cepted accounting principles. The auditor has no reason to give a qualified opinion whichever of the several generally accepted accounting principles are used by the corporation.

EDWARD H. JENNINGS, Ph.D. Michigan 1969. An empirical examination and analysis of the diversification requirements of the common stock investor.

RODNEY D. JOHNSON, Ph.D. State University of New York (Buffalo) 1970. The extinguishment of convertible bonds.

CHARLES P. JONES, Ph.D. North Carolina (Chapel Hill) 1969. The value of quarterly information in predicting future stock price changes.

The purpose of this research is to examine published quarterly information available to an investor at the time of his investment decisions in order to evaluate its usefulness in explaining future short-run price changes and in selecting stocks that will perform well in the short run. The basic hypothesis is that lags have existed in the market in the assimilation of quarterly data, and that investors have not been able to, or did not, adjust instantaneously (or even very quickly) to all financial information.

OKIE M. JOY, Ph.D. North Carolina (Chapel Hill) 1969. The value of limited monthly information in forecasting airline price changes.

ALFRED L. KAHL, JR., Ph.D. Florida 1969. Investment management and the computer: limitations and prospects.

This study presents simulation results of an extended Markowitz-Sharpe portfolio selection model over 3 ten-year performance periods. The investment performance of the model was found to be superior to random portfolio selection, typical mutual funds, and the general market performance at the .01 level of significance for both equal dollars and equal shares buy and hold investment strategies, although neither strategy proved superior to the other.

DONALD KAPLAN, Ph.D. California (Los Angeles) 1969. Corporate capital structure and marginal financing decisions.

ROBERT H. KEELEY, Ph.D. Stanford 1969. Pension plan decisions and corporate financial policy.

ROBERT M. KEITH, Ph.D. Alabama 1969. Guidelines for establishing a marketing cost accounting system in a small-to-medium sized retail business with computer capability.

JAMES C. KINARD, Ph.D. Stanford 1969. The effect of variations in the timing and ordering of presentation of otherwise identical information on expectations.

WILLIAM M. KINCAID, JR., Ph.D. Texas (Austin) 1970. A study of the perception of selected brands of products as foreign or American and attitudes toward such brands.

The basic problem in this study was to investigate the attitudes consumers have toward foreign brands. The results of the study, based on an analysis of ten brands, five foreign and five parallel American brands, showed with varying degrees of conclusiveness the following: that attitudes toward foreign brands not perceived as foreign are not different than attitudes toward American brands, but are different than attitudes toward foreign brands perceived as foreign; and that attitudes toward foreign brands are different with exposure to different amounts of information concerning the brands.

ROGER KLEIN, Ph.D. Michigan State 1969. The personal income tax, dividends, capital gains and the allocation of capital.

This study explores the hypothesis that the current method of taxing corporate income, dividend payments and capital gains leads to a lowering of the cost of capital and induced retention of earnings. Using data from firms in three different industries, a dividend model is tested in order to determine the extent of the tax shelter argument. The performance of firms whose dividend policy is influenced by personal income tax rates are compared with other firms in the same industry. According to the measures of performance used, there was no systematic misallocation of capital due to the tax shelter.

ADRIAN L. KLINE, Ph.D. Michigan State 1970. Implications of responsibility centers for successful implementation of the Department of Defense resource management systems.

On July 1, 1968 the Department of Defense (DOD) officially launched an ambitious program to install a responsibility accounting system which has the potential to revolutionize the management of resources by DOD components. This study established that the system had not attained full benefits generally associated with enlightened responsibility center management. It also developed a research effort which resulted in a set of viable concepts to assist financial managers in achieving these benefits.

LAWRENCE M. LAMONT, Ph.D. Michigan 1970. Technology transfer, innovation and marketing in science-oriented spin-off firms: a conceptual model.

JAMES C. LAMPE, Ph.D. Michigan 1970. A comparison of intuitive deterministic and probabilistic capital budgeting techniques.

MAURICE LANDRY, Ph.D. California (Los Angeles) 1969. Circumstantial variable in accounting for precomputed income of a finance company.

LAWRENCE A. LANG, Ph.D. North Carolina (Chapel Hill) 1969. A heuristic order selection allocation procedure.

RAYMOND L. LARSON, D.B.A. Oregon 1970. A behavioral investigation of the pricing of intra-company transfers and the operation of the industrial firm organized under the profit center concept.

The purpose of this research was to examine, from an accounting and behavioral viewpoint, some of the problems of decentralization and transfer pricing. The hypotheses tested were: that internal accounting must be adjusted to the whole man, not to economic man only, if the accountant is to be successful in solving specific problems of the decentralized firm; that profit center operations lead to optimum performance within the decentralized firm.

V. PARKER LESSIG, Ph.D. Kansas 1970. Numerical taxonomy in market segment identification: a process theory approach.

A theory is developed to be used in the identification of market segments. The theory states that consumers who are similar in their buying behavior and personal characteristics should have similar responses to market stimuli; such a group can be considered a market segment. Numerical taxonomy is used in the identification of these groups. The application of the theory not only adds support to the market segmentation theory but also to process theories in consumer behavior.

RUDOLPH S. LINDBECK, Ph.D. Alabama 1969. An inquiry into the suitability of nonparametric statistical tests for accounting and auditing problems.

ROBERT H. LITZENBERGER, Ph.D. North Carolina (Chapel Hill) 1969. Allocation of corporate capital under uncertainty.

PETER R. LLOYD-DAVIES, Ph.D. Rice 1970. Portfolio selection and the theory of corporation finance.

This thesis is an extension of the Modigliani-Miller theorem under more general assumptions. It starts with a model of portfolio selection, based on expected utility maximization, and derives the conditions for an investor to reoptimize his portfolio in response to a change in leverage. The existence of constrained financial institutions, corporate taxes, costs of bankruptcy, and retained earnings is recognized; each is seen to affect the relationship between leverage and the value of the firm. I conclude that the optimal degree of leverage will be inversely related to the firm's degree of business risk and its rate of growth.

JOSEPH G. LOUDERBACK III, Ph.D. Florida 1970. Ethical considerations in accounting theory formulations.

Accounting writings exhibit various positions relative to ethical considerations in accounting theory formulation. Writings representative of various modes of advocacy were examined from the standpoint of criteria for knowledge. The standards for judgement were the epistemology and related ethical theory of John Dewey. It was concluded that the accounting writings could have drawn their support from three identified epistemologies. The latter were, it was argued, inconsistent with the method of science as presented by Dewey.

LEVIS D. MCCULLERS, Ph.D. Florida 1969. Convertible securities: debt or equity?

In this study a set of decision factors which should be considered in determining the debt or equity nature of a security was developed from accounting, finance, investment, and legal sources. Empirical data relative to convertible securities issued during 1968 were then collected and summarized. The data were evaluated on the basis of the established decision factors. It was concluded that convertible bonds whose conversion value exceed the call price should be classified as equity.

ALBERT C. MCDOWELL, Ph.D. Cornell 1969. A capital market theory with implications for capital budgeting.

RICHARD W. MCENALLY, Ph.D. North Carolina 1969. Seasonal variations in price changes and holding period returns of common stocks.

In an "efficient" market for common stocks nonrandom price movements (with the possible exception of a trend component) should not exceed the costs of exploiting the movements. This study explores the implications of the efficient market model and utilizes a sample of dividend-adjusted holding period returns for 653 common stocks over the years 1945–60 to test for the presence of a seasonal pattern. A number of statistically meaningful seasonal components are found in the series, but for the most part they are not significant compared with the costs of exploitation. The study also investigates the causes of the seasonal movements and their persistence over time.

RICHARD A. MCGARRITY, Ph.D. Northwestern 1969. An application of space transformation in evaluating the urban location pattern of a multi-unit retail firm.

The purpose of this study is to develop a method for identifying and evaluating the spatial patterns formed by retail chains within an urban market. The approach taken (from the point of view of retail management's competitive strategy) involves the transformation of geographic space into a representative demand space using a newly developed technique. As an illustration, the method is applied to the location patterns of supermarket chains in a specific urban market.

JAMES C. MCKEON, Ph.D. Pennsylvania 1969. Conflicting patterns of structural change in wholesaling.

Selling and buying groups are separated in terms of time and space, and their product and service desires are separated in terms of quantity and variety. This time-space-quantity-variety separation is called a discrepancy of assortments and dictates the structure of the channel of distribution. Using the concept of discrepancy of assortments, this dissertation identifies,

measures, and evaluates the structural changes at the wholesale level of thirty-five trades.

JAMES MCKEOWN, Ph.D. Michigan State 1969. An application of a current market value accounting model.

In order to investigate the feasibility of the current market value accounting model developed by R. J. Chambers, two balance sheets and the intervening income statement were prepared according to the Chambers' model to report the activities and financial position of a subject company. A comparison was then made between the accuracy of the revised statements and the verifiability of the conventional statement.

GEORGE W. MCKINNEY III, Ph.D. Stanford 1969. An experimental study of the effects of systematic approaches to strategic planning.

LAWRENCE L. MCNITT, Ph.D. North Carolina (Chapel Hill) 1969. Generating decision networks.

Decision trees may be used to represent multi-stage decision problems. Two concepts are suggested for simplifying the use of decision structures such as decision trees: the decision tree formulation for many problems can be reduced to a much more compact network form; and some problems are of such a well-ordered nature that the computer can be used to both generate and evaluate the decision structure. Two models, a capital budgeting model and an item-by-item sequential sampling model are used to demonstrate the application of these two concepts.

CHARLES D. MCQUILLEN, Ph.D. Florida 1969. Evaluation of regulation A of the Federal Securities Act of 1933.

KARL MAGNUSEN, Ph.D. Wisconsin 1969. Technology and organizational differentiation: a field study of manaufacturing corporations.

ABD-ELHAY A. MARIE, Ph.D. Michigan State 1969. A critical study of asset valuation and income determination under the new uniform UAR (Egypt) accounting system relative to the objectives of economic planning.

In December 1966, a uniform accounting system intended to provide information needed for micro- and macroeconomic planning and control was issued in the UAR to be used by all economic units in the public sector, except banks and insurance. This study examines the underlying theoretical support and rational behind the system, and evaluate the relevance and appropriateness of information provided by it to the intended objectives. Other possible alternatives are examined, developed, and recommended.

CLAUDE R. MARTIN, Ph.D. Columbia 1969. The theories of double jeopardy and natural monopoly and their contribution to marketing as a science.

OTTO B. MARTINSON, JR., D.B.A. George Washington 1969. A standard classification system for the indirect costs of defense contractors in the aircraft industry.

This study examines how indirect (overhead) costs of defense contractors are analyzed within the DOD procurement process. The inadequacy of the present method of viewing indirect costs through overhead rates is demonstrated. Based on a study of five years of costs data from eleven aircraft plants, an indirect cost classification system is developed and used as a framework for constructing a price deflator for indirect costs and an indirect cost model for a plant.

MARVIN MAY, Ph.D. California (Los Angeles) 1969. An investment opportunities stock valuation model based on growth pattern of equity.

JAMES V. MILANO, D.B.A. George Washington 1969. Development and implementation of a management information system.

A study on how to develop a management information system (MIS). The study advocates an evolutionary approach to a MIS which includes a technique of integrating functional data within a modular concept. This technique relies on operational systems analysis, an analytical concept which bridges the gap between data systems analysis and operations research.

GIBBES U. MILLER, Ph.D. Michigan 1970. Accounting and the measure of damages: an analysis of the problem of measuring loss of profits for purposes of obtaining judicial remedy.

MICHAEL L. MURRAY, D.B.A. Oregon 1970. Utility analysis and the selection of an automobile collision insurance deductible.

An application of the von Neumann-Morgenstern utility model to the selection of an automobile collision insurance deductible. Ten persons were each asked one hundred questions concerning their attitudes towards varying dollar values. A polynomial least squares function was fitted to these responses and used in an expected disutility model along with both objective and subjective probability of loss distributions.

GURRAMKONDA M. NAIDU, Ph.D. Michigan State 1969. Systems approach to the marketing aspects of higher education.

The study examines the variables that make educational product package more acceptable to graduate students. Using Michigan State University as a data base a probabilistic and a systems model are used to predict individual purchase behavior and the aggregate demand. The study finds that a) financial aid alone will not attract quality students implying a need for a balanced mix of the educational product package; b) as financial aid per student increases, the incoming quality of students improves and c) as the given budget is spread on many students, the incoming quality decreases.

KAHANDAS N. NANDOLA, Ph.D. Pennsylvania 1970. Conversion of conventional supermarkets to discount operations.

This study analyzes food discounting. It specifically examines the relatively new trend among large supermarket chains to convert some or all of their stores into discount food stores. It tries to explain exactly what happened, why and with what effects. It examines local competition in food retailing and develops a model of an ideal converter.

G. G. NEFFINGER, Ph.D. Maryland 1970. A study of the conceptual and empirical procedures for the allocation of general and administrative costs.

Allocation of all costs is essential for accurate costing and correct price determination (when price is determined by cost). This dissertation examines methods of allocating general and administrative costs to pertinent cost objectives. It recognizes weaknesses in traditional and current allocation methods and provides guidelines for more exact and justifiable allocations by analysis of the literature; survey of current allocation practices; analysis of pertinent legal decisions; by inductive and deductive reasoning to establish and clarify theory.

DAVID W. NYLEN, Ph.D. Florida 1969. An analysis of product-fact versus psychosocial appeals in advertising: an application of the theory and methods of the behavioral sciences to a business problem.

TERENCE O'BRIEN, Ph.D. Columbia 1969. Information sensitivity and the sequence of psychological states in the brand choice process.

ROBERT L. OEHRTMAN, Ph.D. Iowa State 1970. A hierarchical factor analysis of the adjustment problems facing milk bottling firms.

The study determines some of the sociological and psychological values and economic variables which fluid milk bottlers believe to be relevant to the marketing problems that they face. The multiple-group method of factor analysis was used on data collected by a comprehensive questionnaire, and first-order factors were obtained. Then the correlations among first-order factors were factor analyzed by the maximum-likelihood procedure and second-order factors were obtained. These first- and second-order factors were transformed to group and general factors from which hypotheses were generated.

MOONSONG OH, Ph.D. Pennsylvania 1970. The role of international corporations in the transfer of technology to developing countries.

This study is concerned with the significant role international corporations play in transmitting managerial and technical knowledge and skills to developing countries. The study is based on personal interviews with officials of eleven international corporations.

BYRON D. OLIVER, Ph.D. Cornell 1969. Food shopping behavior and attitudes of highrise apartment and suburban consumers.

Four hundred interviews were completed from the highrise apartment population and three hundred-fifty from the suburban population. It was found that highrise consumers were more time oriented, placed more emphasis on convenience and quality and made more use of shopping lists and store advertisements. Suburban consumers found shopping more enjoyable and were more loyal to their favorite supermarket. Both groups indicated they would not object to more automation in supermarkets if it helped to maintain relatively low prices.

LONNIE L. OSTROM, Ph.D. Alabama 1969. An analysis and synthesis of the legal implications of vertical territorial restrictions as a market control device.

JAMES E. PARKER, Ph.D. Michigan State 1969. A study of the predictive significance of several income measures relative to the accounting for extraordinary items and prior period adjustments.

The problem addressed in this study concerns the method of reporting extraordinary items and prior period adjustments in annual financial reports of business corporations. The criterion of predictive assistance was selected as a means of providing a possible solution to the diversity of views surrounding this area of financial reporting. Finally, historical empirical evidence was examined with respect to the predictive assistance criterion in comprising several income reporting alternatives.

LARRY G. POINTER, Ph.D. Florida 1969. A critical examination of the impact of income tax postponements on financial reporting.

Using Standard and Poor's *Compustat,* tax postponement data from 1954–64 were examined for their financial reporting implications. It was inferred that tax allocation procedures should be heavily dependent on future-oriented periodic assessments of probabilities rather than upon past-oriented mechanistic averaging procedures that lump tax postponements together as if their differences in behavior patterns were of no significance.

VERNON E. PONTIUS, Ph.D. California (Los Angeles) 1970. Criteria for determining the change in income reported under generally accepted accounting principles due to changes in the general price level: a predictive model.

ROBERT W. PRATT, Ph.D. Michigan 1970. Time as a dimension of the decision process for household appliances: implications for marketing management.

FRANK R. PROBST, Ph.D. Florida 1969. The utilization of probabilistic controls in a standard cost system.

Probabilities are utilized in the formulation of a variance investigation decision rule. The basic assumption tested was that a probabilistic model could be made operational. Initially four such models were evaluated in connection with several manufacturing

operations. Subsequently, subjective estimates were introduced in an effort to overcome the behavioral problems encountered.

KAVASERRI V. RAMANATHAN, Ph.D. Northwestern 1970. A mathematical programming approach to strategic planning in diversified firms.

This study develops an optimizing model to facilitate the allocation of resources among the dissimilar strategies of a multi-industry firm. It uses the criterion that the return-risk value of the resulting strategic plan to the firm's stockholders must be greater than what they can achieve from alternative investment opportunities in the market. The implications of the proposed model for the firm's capital structure, management control, and information system have been explored.

JOHN RETTENMAYER, Ph.D. California (Los Angeles) 1969. The effect of file ordering on retrieval cost.

JAMES N. RIETH, Ph.D. Purdue 1970. A dynamic synthesis of tactical planning and strategy formulation: an industrial application.

The management planning system simulates a firm's environment by predicting values that market and firm parameters will take and establishes the optimal action plan for management best utilizing available opportunities. Three subsystems constitute the system: The material price forecasting subsystem incorporates a hierarchy of approaches to generate forecasts; the sales forecasting subsystem makes qualified use of subjective judgement; the capability allocation subsystem utilizes linear programming to determine the best distribution of the firm's capabilities.

DAVID E. ROBERTSON, Ph.D. Wisconsin 1970. Managerial transfer practices and internal mobility in large manufacturing firms.

JAMES T. ROBEY, Ph.D. Florida 1969. The economic theory standard for contemporary accounting theory formulation: some implications of changes in the methodology of economics.

The purpose of this work is to study the reliance of accounting inquiry upon methodological principles associated with late nineteenth century economic thought, and to identify in modern economic thought principles offering alternative possibilities for accounting thought. Reliance upon earlier principles has led to a conception of inquiry inconsistent with the contemporary scientific enterprise and hindered the acceptance of new models. The methodological conceptions of modern economics offer promising insights for contemporary accounting thought.

DONALD R. ROGERS, D.B.A. Oregon 1970. Interperiod income tax allocation: a fourteen-year study.

MURRAY R. ROSEMAN, Ph.D. Harvard 1970. Portfolio investment by mutual savings banks.

YEHOSHOA RONEN, Ph.D. Stanford 1970. Some effects of sequential aggregation in accounting on decision-making behavior.

TIMOTHY L. ROSS, Ph.D. Michigan State 1969. The accountant's role in participative decision-making.

The study's findings disclose that increased productivity as a measure of accepting a major organization change in a participative climate of a Scanlon Plan company was significantly related only to attitudes regarding perceived size of economic benefit received and technological change. Additionally, knowledge of the accounting process was significantly related to tenure, position, and two objective measures of participation. Scanlon managers were also less traditional, in many of their attitudes than were non-Scanlon managers.

ROBERT R. ROTHBERG, Ph.D. Pennsylvania 1969. Segmentation in proprietary medicine and toiletry markets: a multivariate analysis.

This investigation was divided into three parts. The first attempted to resolve the different interpretations given to the segmentation concept by marketing practitioners and researchers. The second concerned the correlates of purchasing activity in ten different product categories. In each case, regression analyses were conducted on seven criterion variables using two batteries of predictors. The third compared three methods of numerical taxonomy in the isolation of natural segments in one product market. Hypotheses concerning the distinguishing features of these segments were also tested.

WILFRID F. RYLANDER, Ph.D. Texas (Austin) 1969. Multi-value accounting: new dimensions for price-level adjustments.

Multi-value accounting procedures are new techniques specifically developed for converting historical-dollar amounts into common units of measure possessing instant comparability with similar units from other time periods; providing a simple and easy method for calculating current-dollar equivalents; preserving the original historical-dollar figures; and using these data to facilitate the determination and understanding of monetary gains and losses due to price-level changes.

IL SA-KONG, Ph.D. California (Los Angeles) 1969. Factor market distortions and choice of production techniques.

ISKANDAR SALOUM-HAMWI, Ph.D. Pennsylvania 1970. Life insurance surrender values: concepts, evolution, determination, and practices.

This dissertation analyzes development of nonforfeiture thought and legislation; evaluates statutory methods of determining minimum surrender values; establishes theoretical criteria for equitable surrender values; and examines nature, objectives, reasons, and effects of company practice in granting surrender values.

JOHN C. SCHREINER, Ph.D. California (Los Angeles) 1970. Historically successful portfolio selection rules.

PETER A. SCHULKIN, Ph.D. Harvard 1970. Commercial bank construction lending.

A primary objective of this study was to fill a void in the institutional literature and describe the practices of the major construction lending institution in the United States, the commercial bank. Aggregate construction lending statistics were developed, both for commercial banks and for all institutions. Bank construction lending was examined in the light of bank portfolio theory. The impact of policy on bank construction lending and the impact of construction lending on the volume of construction were also covered.

RANDALL L. SCHULTZ, Ph.D. Northwestern 1970. The development of a marketing planning model through simultaneous equation multiple regression analysis: an airline study.

This dissertation develops a marketing planning model for airline passenger demand in a two-city market. Demand and market share response functions are estimated from empirical data; these structures are then utilized in a normative model of marketing decision-making. Demand is found to be significantly related to personal income, seasonality, and price; market share, to scheduling share, advertising share, and position of the airline in the market (metro shares). The planning model gives profit-maximizing levels for number of flights and dollars of advertising.

MOHAMED H. SELEIM, Ph.D. North Carolina (Chapel Hill) 1969. Technological and human factors as determinants of work-assignment in multi-machine systems with specific application to loom-assignment.

In a number of operating systems (e.g., loom-assignment), different products may be processed on the same machine and the worker may operate simultaneously more than one machine. Three sets of attributes defining K-products, N-machines, and M-workers are investigated and a model for estimating their associated coefficients is developed. The coefficients are incorporated in an optimization algorithm which assigns the N-machines processing the K-products to the M-workers such that the total contribution is approximately maximized.

JAMES H. SELLERS, Ph.D. Arkansas 1970. An investigation of the relative income tax burden of life insurance companies as compared to other corporate organizations.

SHRISH B. SETH, Ph.D. Michigan State 1969. Costing methods for materials processing in the metals service center industry.

This study was an empirical investigation of pre-production processing cost control practices in the metals service center industry. The investigation revealed that a majority of the service centers lack sophistication in this area. These service centers can achieve better control of their pre-production processing costs by using the profit planning and control system suggested in this study. The feasibility of this system has been demonstrated by its successful use by a few service centers.

MICHAEL SIMKOWITZ, Ph.D. New York 1970. Quadratic portfolio selection with theoretically predicted inputs.

DAYA R. SINGH, Ph.D. Michigan State 1969. Investment policy and performance of U.S. subsidiaries in India.

This study is an examination and evaluation of investment policy and performance (profitability and risk) of U.S. subsidiaries in India. These aspects of the U.S. subsidiaries are compared with their parent corporations. The financial statements of the participating subsidiaries were adjusted for the price level changes in India. The study indicates that the subsidiaries' investment policies differed from their parents in many ways. The performance of the Indian subsidiaries is greater than their American parent corporations. Despite higher business risk, the parent corporations were more concerned with the environmental risks in India.

STANLEY R. SINIK, D.B.A George Washington 1969. A method to measure and express contractual plan total cost.

This dissertation is concerned with establishing a method to measure and express the total cost of mutual fund contractual plans in such a manner as to facilitate comparison of such total cost regardless of plan size, time length, or structure and level of rate associated with each component cost.

STANLEY SLOAN, Ph.D. Wisconsin 1969. Management by objectives in a public hospital.

J. EDWARD SMITH, Jr., Ph.D. Purdue 1969. A systems approach to conglomerate corporations.

During recent years numerous multi-product, multi-industry corporations have appeared on the corporate scene. The thesis develops a new approach to the theory of the firm in order to provide a foundation for the formulation of policy statements concerning the conglomerate corporation. The concept of economic balance—the attainment of internal economies through the full utilization of the firm's potentialities and resources—is developed through the use of activity analysis, automata theory, and lattice theory to protray the firm as an interacting system.

RALPH E. SMITH, Ph.D. Kansas 1970. Inflation hedges and common stocks.

EARL W. SNELL, Ph.D. Stanford 1970. A practical approach to the allocation of nonfinancial resources in capital budgeting.

PAUL S. SPECK, Ph.D. Michigan State 1969. The impact of computers on federal income taxation.

This dissertation examines the influence of computers on the federal income tax collecting, tax research, and tax returns preparation activities. The present Internal Revenue Service computer network is described, analyzed and evaluated, and found efficient for the short run but deficient for long-run needs which justifies plans for the installation of an updated system early in the 1970's. Tax research programs have improved and expanded. Eight firms now constitute the newly emerging computerized tax returns preparation industry.

A. EDWARD SPITZ, Ph.D. Kentucky 1969. Some determinants of investing in load or no-load mutual funds: a statistical analysis.

The study investigated the role that cash inflows play with reference to net performance and with reference to the investment company's expense ratio of ten load funds and ten no-load funds from years 1960–67, inclusive. Three simple linear correlation studies and two multiple linear correlation studies were introduced to ascertain the degree of correlation between cash inflows and performance. Further, three simple linear correlation studies were introduced using cash inflows and expense ratios.

RICHARD STAELIN, Ph.D. Michigan 1969. Information seeking and duration of the purchase decision process for new cars and major household appliances.

JOHN D. STOFFELS, Ph.D. Michigan State 1969. The use of margin credit in the trading of securities.

The results of this study support the hypothesis that stock margin trading is speculatively oriented to near-term price changes and that, unlike actions of speculators in classical theory, such trading destabilizes stock prices. Tests of the hypothesis included: development of a demand/supply model explaining changes in aggregate broker-extended margin credit; and analysis of individual margin account transactions evaluating both risk and turnover characteristics of securities and the response of trading to stock price changes.

DAVID R. STRATTON, Ph.D. Texas (Austin) 1970. An analysis of consumer behavior toward pricing as related to education and income groups.

This is a study of consumer purchase preference and brand evaluation for a consumer product. Subjects were asked to select from among six brands of margarine, at their households, for sixteen selection periods. The brands of margarine were divided into two price lines. The participants were asked to evaluate the six brands after the tenth selection and after the sixteenth selection. After the tenth selection, one of the high-price brands was lowered in price. The selection procedures and the evaluation of the various brands available for selection were evaluated by comparing differences between different income and educational groups.

EDWARD M. TAUBER, Ph.D. Cornell 1969. HIT: Heuristic ideation technique, a systematic procedure for new product search, with an experimental application in the food processing industry.

HOWARD E. J. TENNANT, D.B.A. Oregon 1970. Consumer preferences for nonstructural wood materials in new home construction: 1963–64.

This is a study to isolate factors which appear to explain the role played by the home buyer in selection and use of non-structural wood products in construction of new, single family dwellings. Research objectives were to determine specific desires for non-structural wood products, why such preferences exist, the characteristics of consumer groups in the market for these products, and to relate findings to the development of long range marketing stategies.

BOYD J. TODD, Ph.D. North Carolina State 1969. Prediction of hospital utilization with modified exponential smoothing techniques.

Exponential smoothing techniques were used to predict the utilization of specified facilities in a short-term general hospital including medical, nursery, obstetrics, pediatrics, psychiatric, service and surgery. Treated as separate entities were clinic visits, emergency room visits, deliveries and operations. The results indicate that exponential smoothing techniques with an added integral term provide excellent forecasts for series which exhibit trend, seasonal variation and random variability, characteristics that are found in the demand for a wide variety of products and services.

DALE G. WEIGHT, D.B.A. Oregon 1970. The development and application of a demand deposit costing model for small banks in the Pacific Northwest.

BENJAMIN B. WEISMAN, Ph.D. New York 1970. An empirical study of 21 criteria employed in investment analysis.

MARVIN WEISS, Ph.D. New York 1969. An empirical investigation of the revenue maximization hypothesis and of the relationship between functional background and corporate economic performance.

JOSEPH H. WELBORN, Ph.D. Arkansas 1970. An inquiry into the establishment and utilization of standard costs for delivery function of retail stores.

JAMES L. WIEK, Ph.D. Michigan State 1969. An analysis of the relation of vertical integration and selected attitudes and behavioral relationships in competing channel systems.

Relationships between the structure of competing egg industry marketing channel systems and system-member attitudes and behavior were studied with the aid of Q-methodology, other psychometric scales, and multivariate analysis. Significant relations were found between vertical integration and levels of intra-channel cooperation, conflict, innovativeness, modern

norms, channel reference group usage, and marketing orientation. Channel system openness—measured in terms of central coordination, feedback, cohesiveness, and attitudes—appeared to be directly related to competitive effectiveness.

TILTON L. WILLCOX, D.B.A. George Washington 1969. Survival considerations for the small broker-dealer in the District of Columbia.

The securities business is becoming increasingly intolerant of small poorly managed general securities brokerage firms. The dissertation purpose was to establish determinants for survival and growth for small securities broker-dealers in the District of Columbia. This objective was approached through the identification and evaluation of major environmental and managerial problems of small securities firms. Specific firm characteristics and management actions that will enhance survival and growth were determined.

JON R. WILLIAMS, Ph.D. Arkansas 1970. An analysis of retail credit costs of selected retail stores in Arkansas.

GARY YAMASHITA, Ph.D. Columbia 1970. Cash tender offers.

MARTIN E. ZWEIG, Ph.D. Michigan State. An analysis of risk and return on put and call option strategies.

Performance, as measured by return per unit of risk taken, was found on 54 put and call option strategies tested over a seven-year period by means of over 45,000 hypothetical transactions. The results disclosed that option writing was generally an inferior strategy and that some option buying strategies were highly profitable but only at the expense of undue risk. Ordinary stock investing displayed superior return-risk performance than all 54 option strategies, and with the exception of buying six-month calls on more volatile stocks, the differences in performance were all statistically significant.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

ARTHUR J. ALEXANDER, Ph.D. Johns Hopkins 1969. The price guideposts: application and effect.

This study of price guideposts focuses on the policy of public and private confrontations between government and industry from 1965 to 1967. It analyzed forty-six incidents of government intervention. Several factors affecting government influence over specific price increases were examined and found to be statistically significant: concentration ratio, rate of return, markup, confrontation strength, and time. The estimated impact on various price levels of eighteen cases of lowered prices was one to three percent.

LOUIS J. J. ALLAIN, Ph.D. Purdue 1970. Capital investment models of the oil and gas industry: a systems approach.

WILLIAM B. ALLEN, Ph.D. Northwestern 1969. A model of the demand for transportation: the case of air freight.

A general model for transport demand based on the theory of the firm is developed. In order to test the model of modal choice, the competition of air and sea in the North Atlantic was selected. A statistical technique, discriminant analysis, which can separate all or nothing situations like shipment by air or water was developed into a probability model of modal choice. Demand curves for air freight based on percentage changes in rates were then derived.

ROBERT J. ANDERSON, JR., Ph.D. Pennsylvania 1969. Application of engineering analysis of production to econometric models of the firm.

SAMUEL H. BAKER III, Ph.D. Virginia 1970. Empirical tests of the sales maximization hypothesis.

After reviewing previous attempts to test Baumol's sales maximization (SM) hypothesis, a formal statistical test is developed in which both a null hypothesis of SM and an alternative hypothesis suggestive of profit maximization are posed in a clear-cut manner. The test procedure uses time-series data for 30 firms. The results tend rather strongly to reject the SM theory. Also, test results, including new ones, are reported for executive compensation-sales-profit equations.

ALEXANDER BELINFANTE, Ph.D. California (Berkeley) 1969. Technical change in the steam electric power generating industry.

SANFORD, V. BERG, Ph.D. Yale 1970. Structure, behavior, and performance in the scientific journal market.

STANLEY J. BIRKIN, Ph.D. Alabama 1969. Error reduction programs: a factorial investigation of their success in various industries.

MARC BLUM, Ph.D. Columbia 1969. The failing company doctrine.

GORDON F. BOALS, Ph.D. Princeton 1970. An economic analysis of book publishing.

This study attempts to develop a model of the behavior of book publishers for use in evaluating the effects of copyright legislation and in testing hypotheses about the behavior of firms under actual market conditions. It considers both the optimal strategies available to firms and the actual conduct of firms and of the industry as a whole. The analysis is based upon in-depth case studies of two firms, as well as generally available industry sources.

J. F. BURNEY, JR., Ph.D. Alabama 1969. PERT and PERT/COST: a study and evaluation of its industrial utilization in Alabama.

VICKI CARNAHAN, Ph.D. California (Los Angeles). An economic analysis of the U.S. patent grant.

The nature of the property right conferred by the patent grant is delineated through an examination of limitations placed upon exploitation of the patent by antitrust law. The devices utilized by patent holders are analyzed in terms of profit maximizing behavior, in an attempt to determine whether use of each device represents greater exploitation of the monopoly conferred by the patent or extension of monopoly power into new markets.

THOMAS G. COWING, Ph.D. California (Berkeley) 1970. Technical change in steam-electric generation: an engineering approach.

LYNN M. DAFT, Ph.D. Michigan State 1969. Competition in the grocery retailing industry.

LARRY DARBY, Ph.D. Indiana 1969. An evaluation of federal regulation of common motor carriage.

Regulation is evaluated within the framework of the social costs and social benefits attributable to the present scheme. The social dividends from regulation are found to be not quantifiable, but an examination of the data, and the behavior and remarks of shippers suggests the presence of appreciable social benefit traceable to regulation. The costs of regulation include the costs of maintaining and utilizing the regulatory mechanism and the economic costs in terms of foregone efficiency in resource use. The thesis concludes that regulation need not be abandoned—at least for the present—but, that regulation should allow and encourage a cost-oriented rate structure.

DONALD A. DAWSON, Ph.D. Western Ontario 1969. Economies of scale in the secondary education sector in the Province of Ontario.

The thesis tests for the existence of economies of scale in single schools and school boards in the secondary school education sector in the Province of Ontario. In addition, a comparison is made between costs existing in the private and the public portions of the secondary education sector. Schools exhibit constant or declining costs while school boards exhibit areas of both increasing and decreasing costs. Public schools exhibit lower costs than private schools.

STEVE A. DEMAKOPOULOS, Ph.D. New York 1970. Methods and efficacy of long-range industry forecasts: a case study of the domestic air cargo industry.

The study is an examination of long-range forecasting methods and forecasting accuracy in the air cargo industry. The analysis was carried out to determine the characteristics of efficient long-range aggregative air cargo forecasts and to draw conclusions regarding the desirable ingredients of long-range aggregative industry forecasts in general. The main causes of forecast error were isolated, and several absolute and relative tests of forecasting accuracy were employed.

HAIM DORI, Ph.D. New School 1970. Bank size and operating ratios: a study of commercial banking, 1965.

The study concentrates on testing the hypothesis of economies of scale for banks larger than fifty million dollars in total assets. The study is based on a cross-section of one hundred and ten banks selected at random using methods of multiple regression. The results indicate a reduction in the unit cost of operating bank size, after due account is taken of other factors influencing bank costs. An analysis of gross earnings and net earnings per unit of assets is also undertaken.

ALBERT J. ECKSTEIN, Ph.D. Georgetown 1970. Price and wage behavior for the U.S. economy: an inquiry into the value of disaggregation.

This study examines wage, price and productivity behavior of major sub-sectors and manufacturing industries in the United States during the postwar period. Productivity growth is shown to be an important endogenous element in the joint determination of wages and prices, although the empirical results were inconclusive in supporting a guidepost or simple neoclassical-type theory. However, the diversity of behavior of disaggregate units suggests that aggregation characteristics are important to the formulation of a coherent policy for price stability.

DOUGLAS A. ELVERS, Ph.D. Michigan 1969. An investigation into the determination of the optimal frequency for monitoring critical path schedules as they are applied in the construction industry.

FRANK C. EMERSON, Ph.D. Minnesota 1969. The determinants of residential value with special reference to the effects of aircraft nuisance and other environmental features.

A Cobb-Douglas-type "housing services production function" is tested with a sample of existing residences in a predominantly middle or upper income area of Minneapolis, Minnesota. The effects of 25 dwelling and environmental characteristics are investigated. It appears that there are significant external diseconomies associated with proximity to aircraft operations, freeways, and schools. Open green space is a significant source of external economies.

E. DAVID EMERY, Ph.D. Minnesota 1969. An investigation of the potential welfare effects associated with rate of return regulation of the steam-electric industry in the United States.

This study is an investigation of the potential welfare gains and losses that would accompany effective rate of return regulation of the electrical power industry in the United States. The Averch-Johnson model was used to predict the behavior of the regulated utilities and estimates of conventional welfare losses and gains were made for varying degrees of effectiveness for the earnings constraint.

GEORGE ENGLER, Ph.D. California (Los Angeles) 1969. The typewriter industry: the impact of a significant technological innovation.

RICHARD L. ERNST, Ph.D. California (Berkeley) 1970. Capacity adjustments in oligopolistic markets.

JAMES S. FRALICK, Ph.D. Syracuse 1970. A microeconomic analysis of the impact of changes in the depreciation tax credit laws in the investment spending of 26 firms.

CHARLES M. FRANKS, Ph.D. Colorado 1970. A study of technological change in the U.S. nonfarm economy from 1869 to 1960.

WOLFGANG W. FRAZN, Ph.D. Washington State 1970. The timing of the impact of defense contracts and private orders on employment in an aircraft industry, 1958–68.

JO F. FREEMAN, Ph.D. Arizona 1970. Development of the concept of entrepreneurial profit, in the context of a changing environment.

AUGUST B. FRISCIA, Ph.D. New York 1970. Industrial retardation and economic growth.

This is a case study of secular and structural change in the bituminous coal industry of the United States from 1920 to 1960. It is maintained that the decline of this industry provides some empirical support for a restated theory of industrial retardation in mature economies concerned about the maintenance of good economic growth.

JOSEPH C. GALLO, Ph.D. Missouri (Columbia) 1969. Computer programs to detect collusive bidding on public contracts.

Three computer programs were developed to identify patterns in relationships among bidders. These programs permit the examination of market shares on a geographical basis as well as over time. They will also identify peculiar patterns in bids. Such programs cannot conclusively identify the existence of collusion but may permit an economist to testify to the existence of such collusion.

GLENN C. GAMBLES, Ph.D. Maryland 1970. Structural determinants of profit performance in U.S. manufacturing, 1947–67.

VICTOR P. GOLDBERG, Ph.D. Yale 1970. Interproduct competition and market power: a case study of cellophane.

PAUL HAAS, Ph.D. Boston College 1970. The glass container industry: a case study of oligopoly and antitrust.

The glass container industry offers an interesting case study of the effectiveness of antitrust litigation which emphasizes conduct. Prior to 1945 the industry structure was oligopolistic with a patent-pool which restricted entry. High profits and a conscious attempt to restrain progress resulted. Subsequent to the 1945 litigation, the profit rate fell and competition from substitutes increased, suggesting competitive performance, however maintenance of competitive pressure will require close scrutiny of mergers among firms producing substitutes.

JAMES L. HAMILTON, Ph.D. Duke 1969. Regulation and congestion in the markets for flue-cured leaf tobacco.

CHARLES R. HANDY, Ph.D. Cornell 1969. A model of market structure and competition in the food industries.

This research extends the industrial organization model to include bilateral interactions between food manufacturing and retailing. Behavior peculiar to the distribution oligopoly, as well as other subsets of food industry, is identified. The oligopoly manufacturers tend toward product competition while the oligopoly distributors (large food chains) tend toward private label and price competition. Rivalry is effective in providing real consumer alternatives. Performance of one industry cannot be meaningfully appraised independently of the other.

STANLEY HARDY, Ph.D. Michigan State 1969. Vertical and horizontal structures in the copper industry.

This study examines vertical and horizontal structures existing in the copper industry. Ownership, joint membership in two or more vertically integrated groups, contractual relationships, and direct and indirect interlocking directorates are the interconnections forming these structures.

JAMES E. HASKELL, Ph.D. Kansas State 1970. Optimum size and number of farm petroleum distribution facilities: a study of plant and distribution costs.

This study had the objective of determining the equilibrium size of bulk petroleum plants in Kansas. Plant and distribution costs were estimated by economic-engineering techniques and then combined to form long-run average total cost functions. Results of the study point to fewer and larger bulk petroleum plants. There are substantial economies of size in plant operations up to the 1–1½ million gallons per year size but beyond this range marked economies do not exist.

EARL A. HASLETT, Ph.D. Toronto 1970. Factors in the growth and decline of the cheese industry in Ontario, 1864–1924.

A quantitative approach is used to explain the growth and decline of the cheese industry. Secondary data are supplemented by primary data compiled from cheese factory records. An examination is made of changes over time in the six major factors that determined the supply and demand functions. Factors analyzed are changes in: institutions; market demand; cost of transporting milk and manufacturing cheese; regional supply of cheese; technology; and cost structure of farms.

W. Lee Hoskins, Ph.D. California (Los Angeles). Public policy towards bank mergers.

This study examines some of the central issues with regard to public policy towards mergers by focusing on the commercial banking industry. The problem of merger regulation in banking is examined from the perspective of achieving and maintaining an efficient market structure within the constraints of government restrictions on entry into the industry. Emphasis is given to the development of bank merger standards, market definitions, and the use of concentration ratios as policy variables.

Sidney H. Ingerman, Ph.D. California (Berkeley). Industrial growth and wage structure formation in the California penninsula aerospace industry.

Randall H. Hoemke, Ph.D. Michigan State 1969. State branch banking laws: their effect on economic activity.

Annual (1946–66) balance sheet data are summed into statewide, limited, and unit group totals. Portfolio is divided into liquidity, U.S. government bonds, other securities, loans and discounts. Hypothesis, stipulating no difference between systems, is accepted for limited and rejected for statewide. Statewide invests largest proportion locally making most significant contributions to state's economic activity.

David Holmes, Ph.D. California (Los Angeles) The economic nature of the credit union and its role in rural credit reform: a case study of Venezuela.

Lal C. Jagetia, Ph.D. Alabama 1969. A comparative study of administrative ratios and organization size in the textile industries of India and the United States: 1958–67.

Harold Katz, Ph.D. Columbia 1970. Decline of competition in the automobile industry: 1920–40.

This dissertation attempts to determine why competition in the industry declined by an examination of economies of scale. Regression analysis and case studies of the sixteen largest companies show the extent of, and changes in, economies of scale over the years. Then the production and distribution processes are examined to provide a rational explanation for such economies. The results are summarized by synthesizing an average cost curve for the industry and showing its evolution.

Thomas M. Kelly, Ph.D. Oklahoma State. The influences of size and market structure on the research efforts of large multiple-product firms.

Data from a 1950 Federal Trade Commission survey of the 1,000 largest manufacturing firms are used in a multiple regression model to examine hypothesized relationships between several variables and firm research and development intensity. Increased size and market share do not lead to increased R&D effort. Diversified firms tend to be more research intensive than highly specialized firms. There is tentative evidence of a quadratic relation between market concentration and research.

William H. Kelly, Ph.D. Maryland 1970. The influence of market structure on the profit performance of food manufacturing firms.

In this study the firm is viewed as a multimarket enterprise, operating in a more or less unique set of markets, depending on the extent of its diversification. Values of market structure variables reflecting seller concentration, product differentiation, and conditions of entry are computed for each firm as weighted averages of the corresponding values in each of the firm's individual product markets, using firm shipments in these markets as weights. Profitability of food manufacturing firms is found to be significantly related to each of the major variables of market structure, verifying the theory that market structure has a significant influence on the profit performance of firms.

Donald Kingdon, Ph.D. California (Los Angeles) 1969. The management of complexity in a matrix organization: the socio-technical approach to changing organizational behavior.

Benjamin P. Klotz, Ph.D. Minnesota 1969. Scale and substitution in manufacturing plants: a study of selected U.S. industries.

Production functions are fitted to the cross-section data of 1700 plants in 25 4-digit manufacturing industries, mainly the primary metals group. Depending on the production model assumed, alternative estimates are obtained for the degree of returns to scale and for the elasticity of substitution between capital and labor. The constant-returns version of the Cobb-Douglas function is not rejected but the results are statistically imprecise and clouded by a complex of theoretical biases.

Joseph Kushner, Ph.D. Western Ontario 1969. Economies of scale in the general hospital industry.

The primary purpose of this dissertation is to ascertain the extent and magnitude of scale effects in general hospitals in the Province of Ontario. In addition, the importance of other determinants of hospital costs is examined.

Waldo J. Lombardi, Ph.D. Columbia 1970. The changing structure and viability of the Italian footwear industry since the Treaty of Rome.

Contrary to the widely held notion that in the absence of a strong anti-monopoly tradition, the natural form of industrial organization would be the "dominant-firm" type, the findings of this thesis reveal that Italy—a country lacking in such tradition—has not experienced significant merger and related activity producing dominant arrangements. Rather, outside of the state monopolies, the natural arrangement appears one of "symbiosis" manifested by the "mutual coexistence" of small and large firms.

William F. Long, Ph.D. California (Berkeley) 1970. An econometric study of performance in American manufacturing industry.

ROBERT E. LOONEY, Ph.D. California (Davis) 1969. Antitrust and mergers: an economic analysis of some recent court decisions.

WILLIAM A. LOVETT, Ph.D. Michigan State 1969. Patents and headstarts: a study of the polyolefin plastics.

The polyolefin experience tends to confirm the assumption that headstart advantages i.e., lower costs or increased demand arising in favor of innovators, could often be a substitute for strong patent protection. After an initial period of strong patent protection which helped to nurture basic innovations, a compulsory, reasonable royalty licensing decree greatly reduced barriers to entry. Competitive rivalry increased, industrial efficiency was improved, and further innovation resulted. Two public policy lessons are drawn: greater antitrust efforts should be made to reduce excessive patent protection in other industries, and "improvement patents" of a shorter duration—let us say 5 or 6 years, would be desirable to reduce the undue reinforcement of headstart advantages.

JAMES A. MCMAHON, Ph.D. Boston College 1970. Price determination by a cartel: a study of the international air transport association's pricing of the economy class in the North Atlantic.

The cartel model used in this dissertation states that differences among member firms with respect to what the cartel price should be are related to differences in their unit production costs. A cost model is developed to generate unit cost estimates. These estimates are compared with the rating positions on price of member firms. The prediction of the cartel model is found consistent with the sample data.

RICHARD J. MARASCO, Ph.D. California (Berkeley) 1970. The organization of the California tuna industry: an economic analysis of the relations between market performance and conservation in the fisheries.

This study is an evaluation of the California tuna fishing industry. Specifically, attention is focused on "Yellowfin," the most valuable species to the California tuna fishermen. Examined and evaluated are the forces influencing the behavior of the individual fisherman and consumer.

ROBERT T. MASSON, Ph.D. California (Berkeley) 1969. Executive compensation and firm performance.

MARIN MAYDON-GARZA, Ph.D. Massachusetts Institute of Technology 1969. Studies in technological change as an investment process: theoretical and empirical aspects.

PETER C. MAYER, Ph.D. California (Berkeley) 1969. Technical change in the typesetting of daily newspapers.

MADHAVI P. MEHTA, Ph.D. Florida 1969. Prices and competition in the pharmaceutical industry.

LEONARD MEREWITZ, Ph.D. California (Berkeley) 1969. The production function in the public sector: the production of postal services in the U.S. post office.

JAMES C. MILLER III, Ph.D. Virginia 1969. Scheduling and airline efficiency.

CARLISLE E. MOODY, JR., Ph.D. Connecticut 1970. Allocative efficiency of U.S. machinery manufactures.

This is a pilot study using econometric techniques to construct and apply tests for the existence of misallocation of resources in the manufacturing sector of the U.S. economy. Based on the Cobb-Douglas production function estimated according to analysis of covariance regression techniques, the study analyzes the allocation of resources in these sample industries according to the traditional static welfare model and a dynamic welfare model. The study incorporates a discussion of the price of capital to the firm, a simple measure of the flow of capital input into production, and a discussion of the problems of dynamic welfare analysis.

CHESTER W. NEEL, Ph.D. Alabama 1969. Network models: an inquiry of their feasibility and acceptance as management tools on industrial construction projects.

JOSEPHINE E. OLSON, Ph.D. Brown 1969. Discrimination in motor carrier class rates: the effects of regulation.

This dissertation examines the structure of class rates and minimum charges for motor common carriers in two regions of the United States. It tests whether average costs or monopoly price discrimination can explain the rate structure. It concludes that recent New England rates are generally explained by costs (with a considerable markup) and Middle Atlantic rates by the monopoly price discrimination hypothesis.

STANLEY ORNSTEIN, Ph.D. California (Los Angeles). The relationship of market structure to the profitability of leading firms in select U.S. industries, 1950–60.

DALE ORR, Ph.D. Northwestern 1970. A measure of capital services in the U.S. manufacturing industries.

This study defines and employs a measure of capital services which is devised specifically for the empirical analysis of production. The proposed measure of capital services is based on the consumption by capital of a variable input; specifically, electricity consumed for production purposes. For most specifications of either short- or long-run economic behavior the capital services series defined and constructed in this study performed better than the conventional measure in the empirical analysis of production.

JERROLD M. PETERSON, Ph.D. Illinois (Urbana) 1969. Product quality and technological change.

RAYMOND PICCINI, Ph.D. Columbia 1970. An analysis of the merger activity of large industrial firms: 1948–65.

LESLIE K. POMEROY, JR., Ph.D. American 1970. An evaluation in terms of the public interest of the civil aeronautics board's route strengthening policy for local service air carriers.

In fiscal year 1966, the Civil Aeronautics Board adopted a policy of certificating the local service air carriers to provide air transportation services in the high density, competitive markets served by the trunk air carriers. The dissertation analyzes this policy with respect to its impact on direct subsidy and air transportation services to small communities. The conclusion is reached that the route strengthening policy is not in the public interest.

PHILIP N. REEVES, D.B.A. George Washington 1970. A study of adoption of electronic data processing in the hospital industry.

This study extrapolates from a sample to develop a national profile of hospitals which use EDP. Voluntary short-term hospitals in urban areas are the predominant users. Most EDP systems are simple, single purpose applications. Using adoption of EDP as an indicator of innovation, this study shows that findings of prior research on health department characteristics associated with innovation also are valid for the hospital segment of the health-care industry.

RALPH M. ROBERTS, JR., Ph.D. Alabama 1969. A dynamic programming model for executive career analysis with a case study of the aerospace industry.

JEFFREY H. ROHLFS, Ph.D. Massachusetts Institute of Technology 1969. Economic analysis of the mercury industry.

LOUIS A. ROSE, Ph.D. California (Los Angeles). Monopoly rents of VHF television stations: a study in industry and regulatory commission behavior.

Do VHF television stations obtain monopoly rent? FCC licensing practices, applicant queues, etc., are consistent with legally restrained entry of new stations. National Association of Broadcasters' promulgation and attempted enforcement of industry program and commercial standards are consistent with collusion to suppress competition among established stations. Station intangible property values are consistent with monopoly rent averaging $7,000,000. There are alternative explanations for all these observations; but no observations are inconsistent with the monopoly rent hypothesis.

M. JAMES SALVATE, Ph.D. Columbia 1970. Project management in the construction industry.

NATHAN E. SAVIN, Ph.D. California (Berkeley) 1969. Non-homogeneous inputs of the multi-profit firm: a case study in lumber output forecasting.

CHRISTOPHER SCHAEFER, Ph.D. Fletcher School 1969. The American technology challenge and European integration: a response analysis.

RAYMOND C. SCHEPPACH, Ph.D. Connecticut 1970. A Canadian-U.S. productivity comparison.

RICHARD L. SCHMALENSEE, Ph.D. Massachusetts Institute of Technology 1969. On the economics of advertising.

JEROME E. SCHNEE, Ph.D. Pennsylvania 1970. Research and technological change in the ethical pharmaceutical indusery.

This thesis consists of three sets of empirical studies, all of which are designed to promote a better understanding of industrial research and development. The first two studies deal with the determinants of development cost and the accuracy of development cost and time estimates within one ethical pharmaceutical firm. The third study investigates the relationship between pharmaceutical innovation and firm size, the sources of pharmaceutical innovations, and the time interval between pharmaceutical discovery and innovation.

ALAN M. SILVERMAN, Ph.D. Columbia 1969. Management behavior and textile industry fluctuations.

This study relates changed management behavior to the lessening of fluctuations in production and prices of textile products. Production and purchasing procedures are examined, incorporating interviews with fabric manufacturers, cloth converters, garment producers, and store buyers. Price expectations, particularly of converters, are stressed. With improved decision-making methods and measurements adopted by more firms, adjustments of fabrics production to changes in demand have become more rapid and accurate.

AJIT SINGH, Ph.D. California (Berkeley) 1970. A study of postwar take-overs in U.K. and a contribution towards a theory of take-over bids.

LOWELL C. SMITH, Ph.D. Alabama 1969. An evaluation of public policy proposals for the regulation of private pension plan eligibility requirements and vesting provisions.

MICHAEL SZENBERG, Ph.D. City University of New York 1970. The economics of the Israel diamond industry.

SIROUSSE TABRIZTCHI, Ph.D. Columbia 1970. Spatial economics in manufacturing in Iran.

WILLIAM B. TYE, Ph.D. Harvard 1970. Economic costs of the urban mass transportation capital grant program.

To test the undercapitalization assumption implicit in the urban mass transportation capital grant program of the U.S. Department of Transportation, this dissertation develops a model of optimum motor bus

replacement. The data from Cleveland and Chicago show the undercapitalization hypothesis to be totally without empirical support and that substantial inefficiencies are inherent in a subsidy to capital alone. As an alternative, it is recommended that federal grants to local transit operations equal a percentage of local government contribution calculated on a cost accrual basis.

JACK S. VENTURA, Ph.D. Georgetown 1970. Railroad freight cars: economic models of investment and allocation.

After a linear programming treatment, the problem of freight car distribution is integrated into a more general theory of the rail firm. An algebraic example of freight car stock determination traces the influence of detention time off line and the freight car rental (per diem) rate. Investment models are estimated, using series of real net investment, deflated per diem rates, traffic, and other variables.

ROBERT L. VIETS, Ph.D. Texas (Austin) 1970. Financial aspects of the class I, group C, motor carriers of general commodities since 1957.

An analysis of the financial positions and financing practices of the motor carriers for a ten-year period. A determination of the degree of risk of balance sheet account fluctuations and the resulting profitability of the firms measured by return on assets, common equity, and revenues. Development of a probabilistic approach to debt capacity, establishing minimum, maximum, and most favorable levels of debt.

STANLEY L. WARNER, Ph.D. Harvard 1970. Innovation and research in the automobile tire and tire supporting industries.

The propensity to innovate in the automobile tire industry is examined in terms of industry structure, buyer-supplier relationships, and the costs and risks of managing a research program. Fifty-three major passenger tire innovations, occurring between 1910 and 1965, are identified and become the basis for delineating the structure, time, and risk dimensions of the innovation process. The study concludes with an estimate of the long-run social returns to research.

LEONARD WAVERMAN, Ph.D. Massachusetts Institute of Technology 1969. North American natural gas flows.

DAVID A. WILTON, Ph.D. Massachusetts Institute of Technology 1969. An econometric model of the Canadian automobile manufacturing industry.

This thesis analyzes the economic behavior of the Canadian automobile manufacturing industry employing a system of twenty-two simultaneous equations. Major industry variables include retail sales, imports, domestic output, prices, wages, employment, and investment. Control simulations examine the dynamic properties of this annual model over the postwar sample period, and the thesis concludes with estimates for the economic effects of the Canada-United States Automobile Agreement on the Canadian industry.

DONALD F. WOOD, Ph.D. Harvard 1970. Statewide general aviation airport planning: an application of benefit-cost analysis.

Improvements (or initial construction) at about 100 different general aviation airport sites are scheduled for construction by use of benefit-cost analysis. Benefits result from savings in passengers' travel time, and the savings depend upon the length of trip and type of airplane used. Airport projects are interrelated because of travel between them. Costs are those of acquiring land for, and building—or expanding—each airport.

Agriculture and Natural Resources

ROBERT W. ACTON, Ph.D. Purdue 1970. Computer assistance for farm investment decisions.

The subject of this study is the deceptively simple question of "How should a farmer plan the growth of his farm firm when growth implies investment?" Both theoretical and practical issues of using mathematical programming techniques are presented with special consideration given to the "lumpiness" problem characterizing most investment alternatives. Analysis revealed that the crop and livestock programs change appreciably when the investment in capital is integerized and that continuous linear programming approximations to the investment decisions tend to overstate the dollar return on equity capital.

KAMPHOL ADULAVIDHAYA, Ph.D. Purdue 1970. Interregional competition in agricultural production in Thailand.

Estimates were made of the following: the optimal location of crop production and pattern of land resources use; the optimal distribution of four products —rice, corn, cassava, and kenaf—among regions; and the competitive position of each region in producing these crops. A linear programming technique was employed to obtain the optimum solution with respect to minimization of cost of producing and transporting the four products subject to land resource and demand constraints.

JOHN E. AMBROSIUS, Ph.D. Purdue 1970. Optimal tractor and combine replacement policies on corn farms under alternative form growth patterns.

The objective of this study is to define optimal replacement patterns, optimal machinery combinations and optimal machinery investment levels. The results indicate that timing of tractor and combine replacement closely parallels land acquisition. The optimal solution can be highly insensitive to timing of replacement. On the other hand, the optimal solution was highly sensitive to increasing labor costs.

JUAN A. AGUIRRE, Ph.D. Cornell 1969. The economics of milk and beef production in the humid tropics: a case study of San Carlos County, Costa Rica.

The dairy and beef production industry on the northern slope of Cordillera Vulcanica, and on the lowlands at the foot of this slope, in San Carlos County, Costa Rica is described in terms of farm business sizes, rates of production, production practices, capital investments, recent business changes, plans for further development, etc. Differences in farm business success are related to internal farm business characteristics, land characteristics, the nature and growth of agribusiness activities, and government programs.

LEE G. ANDERSON, Ph.D. Washington 1970. The economics of weather forecasting.

Existing methodology of measuring the value of weather forecasting systems is extended so as to include long-term systems. Specifically a means of using new information received during the period of analysis is derived. Geometrical means of analysis are refined and expended. This analysis is then used to discuss the meaning of an improved forecasting system. Finally the new theory is used in several empirical studies.

CHRIS O. ANDREW, Ph.D. Michigan State 1969. Improving performance of the production distribution system for potatoes in Colombia.

DONALD AULT, Ph.D. Missouri 1969. An econometric investigation of price and quantity relationships of egg sizes.

LAKEW BIRKE, Ph.D. Ohio State 1970. The impact of the Dogali irrigation project on the national income of Ethiopia.

ARTHUR F. BORDEAUX, Ph.D. Michigan State 1969. Resource productivity and returns in apple and cherry production on Michigan farms.

GEORGE BRINKMAN, Ph.D. Michigan State 1969. Proposed public investments in agricultural education, infrastructure and production in Nigeria, 1969–1985.

REECE E. BROWN, JR., Ph.D. Purdue 1970. Analysis of the impact of alternative tillage-planting systems on optimal farm organization and income.

A total systems linear programming model was used to determine the effect of alternate tillage-planting systems on optimal farm organization and income. Performance functions were used to determine the effects of resource availability levels on the dependent variables. Optimal solutions showed that all the reduced tillage systems except wheel track were more profitable than the conventional system. The no-till system resulted in the highest profit and highest percentage of land in corn. Conversely, the wheel track system resulted in the lowest profit and lowest percentage of land in corn.

GEORGE W. M. BULLION, Ph.D. Purdue 1970. Estimation of regional demand elasticities for whole milk, United States, 1962–63, 1966, 1961–68.

A broad range or retail price elasticities of demand for whole milk have been derived in studies of consumers located in different areas of the United States. The United States was delineated into regions using a Factor Analysis model to derive composite factor scores for the variables representing the per capita consumption of fluid milk in each state. Twelve regions were defined. The price elasticity of demand for whole milk in the United Seates was found to be within the range of —.15 to —.30. The demographic variables which were significant in explaining the variation in household consumption of milk were race, presence of children, and household size.

JOHN P. CAREW, Ph.D. California (Berkeley) 1969. Economic evaluation of water quality: an application of linear programming.

WILLIAM O. CHAMPNEY, Ph.D. Kansas State 1969. The economics of market turkey production and specialized brooding of poults in Kansas.

This study examined in detail synthesized economies of scale of market turkey production from existing recommended practices. Effects of various poult rearing practices, production programs, and poult prices were included. Sources and magnitudes of economies of scale were identified and compared. Shown also in tabular form are per unit costs of production variables at selected levels which can be employed by industry for comparative purposes.

BOB DAVIS, Ph.D. North Carolina State 1970. An economic analysis of labor use for alternative flue-cured tobacco harvesting and curing systems.

Alternative harvesting-curing systems were analyzed to determine profitability by firm size under varying wage conditions. The effects of adoption of mechanical systems on labor use were estimated, and estimates of the time period required for adoption under alternative conditions were developed. The estimated rate of increase in farm size required for profitable mechanization and the associated reduction in labor use were consistent with recent rates of growth in farm size and rates of outmigration.

KENNETH DEHAVEN, Ph.D. Missouri 1969. Systems of variable rates for milk collection routes: derivation and evaluation.

JACOB DE ROOY, Ph.D. Rutgers 1969. The industrial demand for water resources, an econometric analysis.

A derived demand model is constructed, disaggregating water inputs by function: cooling, processing, power generation, and sanitation. Data are from the chemical industry in New Jersey. Elasticity estimates demonstrates that pollution abatement programs which affect private user costs have a significant impact on patterns of use within the firm. Output elasticity and technological change are also examined. Finally, application of the model is made to projection of future needs.

WILLIAM D. DOBSON, Ph.D. Purdue 1969. An analysis of alternative price structures and intermarket competition in federal order markets.

The primary objective of the study was to examine alternative Class I price structures for federal order markets and to find relationships useful for more efficient pricing and marketing of milk. A model which allocates both bulk and packaged milk products entirely on the basis of costs was employed to predict the long-term economic pressures which may develop under alternative price structures. Plausible forecasts were obtained of the impact of the higher prices on consumption, production inter-market product flows and variables relating to marketing efficiency such as transportation and processing costs.

JOHN P. DUNN, Ph.D. Arkansas 1970. Current municipal water and waste-disposal, pricing practices, and problems in northwest Arkansas.

RICHARD D. DUVICK, Ph.D. Michigan State 1970. Alternative methods of financing growth on Michigan dairy farms.

WILLIAM F. EDWARDS, Ph.D. Florida 1969. Economic externalities in the agricultural use of pesticides and an evaluation of alternative policies.

This study presents a methodology for evaluating total benefits and costs from agricultural pesticide usage. The model employs a measure of welfare consisting of consumers' plus producers' surplus, modified for observable externalities neglected in the surplus calculation. For each of two specified pesticide usage policies, the model maximized this measure of welfare over production of the eight major crops in Dade County, Florida. It was demonstrated that the model could conceptually accommodate environmental constraints connecting levels of pesticide usage with level of residues in the environment.

VERNON L. ENGBERG, Ph.D. Texas (Austin) 1970. Agricultural productivity and economic development in Mexico.

A brief survey of the growth of the Mexican economy precedes an empirical study of the agricultural sector. Increases in agricultural output are attributed to additional inputs of the factors of production and to increases in productivity amounting to a "green revolution" produced by new, high-yielding seed, irrigation, fertilizer, pesticides, and modern cultivation methods. The interdependence between agricultural growth and economic development is demonstrated and is examined in the light of economic development theory.

YILMAZ ESENSOY, Ph.D. Ohio State 1969. Domestic and export demand for U.S. cigarette tobaccos.

BERKWOOD M. FARMER, Ph.D. North Carolina State 1970. Man-hour productivity and future U.S. agricultural adjustment.

Information on advanced production practices was used to generate estimates of potential changes in output and labor input levels for U.S. agriculture between 1964 and 1980. These estimates were adjusted downward somewhat to account for farmer behavior in adopting new techniques. Shifts in demand for products were estimated and contrasted to projected increases in supply. For all commodities except milk, projected changes in demand exceeded projected changes in supply.

WALTER L. FISHEL, Ph.D. North Carolina State 1970. Uncertainty in public research administration and scientists' subjective probability estimates about changing the state of knowledge.

A cost-benefit analysis information system applicable to public research organization was developed and experimentally applied to soybean research conducted by publicly supported agricultural experiment stations. Alternative sources of information on public costs, returns, and variance of the estimates of returns were investigated. Various models were used to aggregate the information so that competing projects could be meaningfully contrasted by a research administrator.

WILLIAM W. GALLIMORE, Ph.D. North Carolina State 1969. Regional differences in growth of the U.S. turkey industry.

Growth of the U.S. turkey industry from 1955–67 was rapid but not uniform among regions. Profitability of turkeys, and broilers, industrial wage rates, turkeys raised per farm, farm wage rates and length of growing season were not found to be correlated with changing regional shares. On the other hand, changes in regional growth patterns were found to be related to industry organization. Following the low prices of 1961, large integrated firms continued to expand, resulting in more rapid growth in the South Atlantic and South Central regions.

JAMES S. GOULD, Ph.D. Cornell 1969. Factors influencing daily wholesale egg quotations in the New York market.

Models for the major sizes of eggs specified that prices were a function of per capital retail movement of each size, storage holdings in New York, military purchases, deliveries to breakers, and commercial movements of each size and competing sizes of eggs. The coefficients of determination (R^2) obtained were good considering the complexities of the daily market, ranging from .77 for pullet eggs to .80 for medium and large eggs.

SANFORD L. GRAY, Ph.D. Washington State 1970. The socioeconomic effects of interregional water transfers upon the state of Washington.

A quantitative statement of the possible effects upon the economic welfare of the state should water transfers occur. Input-output and linear programming models were used to estimate changes in productive potentials of the state, corresponding changes in factor demands, and subsequent effects on state income

and value-added stemming from a change in the water resource base. Guides were developed for state decision makers to use in considering the marginal effects of water diversions.

JERALD A. GUNNELSON, Ph.D. Purdue 1970. A study of the imact of organizational changes in agricultural commodity markets on futures markets.

This study analyzed futures market participation of three large soybean processing firms during the 1957–66 period. The firms used soybean futures contracts to adjust speculative risks and returns of their portfolios by increasing or decreasing their net commodity positions, their net storage service positions, and/or their net processing service positions. Larger relative holdings by these firms, in particular, contracts by date of maturity were associated with greater price variability in those contracts.

ERNEST E. HARDY, Ph.D. Cornell 1969. The agricultural regions of New York State: a mid-century description.

A description of the farm land resources, non-land farm real estate investments, and commercial farm businesses in the 14 major agricultural regions of New York State. Recent agricultural trends are analyzed and opportunities for agricultural development appraised.

STEPHEN B. HARSH, Ph.D. Cornell 1969. Design and application of a computerized farm cost accounting system.

To enable economic analysis of farm business physical and financial data, and to use the rapid computation and vast storage and recall capacities of modern computers, a system of accounting has been devised and programmed for computers for the study of enterprise activities within a business including off-farm transactions. The system provides for the payment of all factors of production at cash or opportunity cost rates, sorts the costs into fixed and variable components, enables the computation of measures of efficiency of production, and provides for estimating the costs and returns from alternative uses of resources.

RAGAA HASSAN, Ph.D. Michigan State 1969. An analysis of the demand for food in Egypt.

ROBERT W. HERDT, Ph.D. Minnesota 1969. An analysis of the aggregate supply function of agriculture in the Punjab (India).

The hypothesis that the aggregate agricultural supply function in the Punjab is positively responsive to changes in relative agricultural-nonagricultural prices is tested by examining data on the eleven most important crops in twelve districts of the Punjab for 1907–46 and 1951–64. The effect of a given percent increase in the price of all agricultural commodities on each crop, holding all other factors constant, is calculated. The resulting changes are aggregated across crops and the aggregate change compared to the base period aggregate output to obtain an estimate of the aggregate supply elasticity for each district.

OMOTUNDE JOHNSON, Ph.D. California (Los Angeles). Economic analysis and the structure of land rights in the Sierra Leone Provinces.

This study describes and analyzes the forms of property rights in land and of land leases in the Sierra Leone Provinces. Theoretical explanations are given for the fragmentation of holdings, the tendency towards individualization of ownership, and the placing of the liability for stronger tenant damages on landlords. The allocative effects of legal and tenure uncertainties in the system and of the legal restrictions on the sale of land to non-Provincials are explained and evaluated.

JOHN IKERD, Ph.D. Missouri 1970. Consumer acceptance and the image of pork: a factor analysis of consumer attitudes.

JOHN M. JORDAN, Ph.D. Purdue 1970. Economic rationality flows: an application of the linear programming transportation model to the U.S. domestic and export corn flows by quarter for 1964.

RONALD E. KALDENBERG, Ph.D. Minnesota 1969. Economic analysis of the optimal size and location of southern Minnesota country elevators.

The economies of size in country grain elevator operation and the diseconomies of size in local grain assembly are analyzed. The Pareto optimal elevator size and market area are estimated with a marginal analysis based on the marginal cost and marginal revenue to an individual patron. The study includes internal plant costs, grain assembly costs, and the Pareto optimal firm size for different grain densities and different capacity delivery vehicles.

ABDOLAMIR KHALILI, Ph.D. Missouri 1969. Optimal economic management of wildlife over time with special reference to Canada geese of the Swan Lake National Wildlife Refuge.

ALAN P. KLEINMAN, Ph.D. Iowa State 1969. The production function and the imputation of the economic value of irrigation water.

This study presents a generalized framework of the production function for crops using irrigation water and considers alternative procedures for imputing the economic value of irrigation water. Although the true response of crops to irrigation water is not known with any degree of certainty, examination is made of the basic soil moisture-plant growth relationships. A conceptual framework of the actual water response function is given with suggestions for needs of additional research. Empirical data from an intensive irrigation fertilizer experiment is fitted to the quadratic equation and static water demand curves are derived and discussed.

MARVIN E. KONYHA, Ph.D. Michigan State 1970. Needs and potentials for escape from poverty

through retraining for families in the Eastern Corn Belt.

JOHN P. KUEHN, Ph.D. Tennessee 1969. Costs and efficiencies of model meat packing plants in the Tennessee Valley.

An economic engineering cost analysis of varying size model beef and swine slaughtering plants to determine economies of size. Costs were examined under differing plant configurations of beef and/or hog slaughtering and processing with varying levels of capacity utilization. Prices were estimated and net revenues were determined for varying degrees of plant utilization.

GORDON L. LANGFORD, Ph.D. Montana State 1969. An economic study of the sheep and lamb industry in the United States, 1930 to 1967.

DONALD W. LYBECKER, Ph.D. Iowa State 1970. Optimum resource use in irrigated agriculture: Comarca Lagunera, Mexico.

This linear programming study suggests that a cotton monoculture should be maintained for current prices and resource supplies under three power technologies. Water and capital are the most critical resources and have MVP's of about $40 (pesos)/hacm and $0.23/peso, respectively. The only livestock activities considered are dairy and they earn about 65 percent return on investment. Required resource bases for selected *ejidatario* income levels are also computed.

HENRY N. McCARL, Ph.D. Pennsylvania State 1969. The mineral aggregate industry in the vicinity of Baltimore, Maryland.

HUGH J. McDONALD, Ph.D. Ohio State 1969. A linear programming model to optimize the transfer cost and facility requirement for U.S. grain exports.

STEPHEN E. McGAUGHEY, Ph.D. Iowa State 1969. Investment criteria for the evaluation and planning of water resource development projects: Peru.

Eleven representative irrigation projects are ranked by the benefit-cost ratio, the social marginal productivity of investment, the output-investment ratio, the labor-investment ratio, the foreign exchange earnings-investment ratio and the internal rate of return. All criteria are reformulated to include present values. A sensitivity analysis is presented of the benefit-cost ratio to a range of shadow prices. Project rankings are derived based upon different preference weights on the national economic objectives.

VASANT M. MEHTA, Ph.D. North Carolina State 1970. India's position in the world peanut and peanut oil markets.

Trade patterns for peanuts and peanut oil in the years 1955 and 1965 were established for major trading nations. Observed trade flows were compared with optimal patterns as found in minimum transport cost models. Major changes between these two years occurred because of the withdrawal of India as a major exporting nation. Preliminary demand and supply projections were made for the year 1975.

NEIL B. MILLS, Ph.D. Ohio State 1970. Production functions of small commercial farms in the East North Central States, 1966.

KENT W. OLSON, Ph.D. Oregon 1969. Expenditure criteria for the development of the national forest road system.

JEROLD F. PITTMAN, Ph.D. North Carolina State 1969. Use of capital budgeting to evaluate selected investment alternatives in the processing of southern vegetables.

Investment value was used as the measure of worth of alternative investments in canning plants designed to handle ten selected vegetables at five selected rates of output in either #10 or #303 cans. When the capital budget was small, the firm was often forced to select smaller and less productive sets of activities due to inability to finance the larger projects. High interest rates also restricted investment selection, even in situations with unlimited capital availability. The quantities of raw products presently available in South Carolina were found to be inadequate to support many of the more productive activities.

BILLY P. PREBBLE, Ph.D. Kentucky 1969. Patterns of land use change around a large reservoir.

This study examines the spatial patterns of land use change surrounding a reservoir. The general hypothesis states that the spatial patterns of land use changes are influenced by economic and geographic characteristics of the reservoir area. Comparisons using data such as slopes, water frontage, and land use changes indicate patterns of land use change surround the lake with factors such as road access, slope, view and location proving to be significantly associated with these patterns.

DONN A. REIMUND, Ph.D. North Carolina State 1969. The influence of grade price differentials on sheller demand for Virginia-type farmers' stock peanuts.

Kernel grade price differentials in effect in 1964 were found to perform well in moving peanuts into commercial channels and in minimizing the cost of operating the peanut price support program. However, programming analysis indicated that grade price differentials underpriced both extra large kernels and other kernels in relation to sound mature kernels. Raising the price of other kernels from 7–10 cents per pound and the extra large kernel premium from 2.5 to 5 cents per pound brought the demand for each kernel grade more closely in line with the available supply.

ABDEL-HAMID Y. SAAD, Ph.D. Cornell 1969. Econometric projections of vegetable crops: New York

State, the United States, and other selected regions. The study involved the estimation of regional supply response functions for 16 different vegetables for the major U.S. production regions; the estimation of a U.S. demand function for each of the vegetables and a derived demand for each region for each vegetable. The functions were used to project production in each region for 1985.

VICTOR SALANDINI, Ph.D. Catholic 1969. The short-run socioeconomic effects of the termination of Public Law 78 on the California farm labor market for 1965–67.

RICHARD K. SCHAEFER, Ph.D. Colorado State 1969. A behavioral derived demand approach to outdoor recreation supply decisions.

The theoretical approach is based on consumer behavioral concepts and on standard economic indifference and production analysis which are used to determine "derived demands" for outdoor recreation resources. The relative value of each resource in adding to the benefits of a recreation experience is determined through the use of multiple regression techniques.

JOHN R. SCHLENDER, Ph.D. Purdue 1970. An information system for financial management of the farm business.

This study was conducted to determine data needed for an adequate farm financial management information system; and to develop an account system. The system developed is capable of processing farm accounts in varying degrees of detail. Enterprise reports can be generated. The system provides reports that are useful in annual financial planning, in planning of loan requirements, and in planing of capital acquisitions.

ALVIN SCHUPP, Ph.D. Missouri 1969. An economic evaluation of sire and length of feeding in an acceptance of beef loin steaks and on pricing accuracy in the beef marketing system.

SALEM V. SETHURAMAN, Ph.D. Chicago 1970. Long-run demand for draft animals in Indian agriculture.

Long-run parameters are estimated using the cross-sectional and inter-temporal variations in data. The model recognizes the different degree of variability of the inputs as well as the adjustment in input levels resulting from the introduction of new inputs. The value of this input relative to all input services will decline in the future through the introduction of new inputs rather than as a result of changes in the employment of traditional inputs.

VISHNOO P. SHUKLA, Ph.D. Cornell 1970. An economic analysis of resource use in farming, Jabalpur District, Madhya Pradesh, India, 1967–68.

This study of 172 Indian farms points out policy issues, resources to be expanded, and policies to be pursued to facilitate modernization of agriculture. Programming indicated increased irrigation and advanced technology could increase total farm incomes 70 percent. Action needed includes accelerated research, education of a variety of technical persons, and an extension program that will expedite farmer adoption of research findings.

SIMON M. SIMON, Ph.D. New York 1969. Economic legislation of taxation: a case study of depletion in oil and gas.

LYNN A. STANTON, Ph.D. Louisiana State 1970. Comparison of selected linear programming techniques for planning farms using Louisiana farm business analysis records.

DONALD G. STITTS, Ph.D. Minnesota 1969. Price efficiency in selected federal order milk markets.

This study's objective was to evaluate milk pricing under the federal order pricing program. A general normative solution was designed so that prices within and between markets would be efficient in respect to the time, place, and form criteria. Upon comparing the empirical data of the six selected federal milk market orders and their respective normative solutions it was concluded that administered prices in all federal order markets did not meet the efficiency criteria.

ROGER P. STRICKLAND, Ph.D. Michigan State 1970. Combining simulation and linear programming in studying farm firm growth.

NORMAN P. SWENSON, Ph.D. Washington (St. Louis) 1970. An economic analysis of federal timber marketing policies in southeast Alaska.

The study evaluated the U.S. Forest Service's policy of prohibiting the export of unprocessed timber from the national forests of Alaska. A vonThunen land rent model was used to compute stumpage residuals associated with processing the timber in Alaska, Puget Sound, and Japan. The highest stumpage residual, $26.20 per thousand board feet, would be obtained if the timber were sold to the Japanese in the form of logs. The study concluded that the present nonexport policy should be replaced with a limited log export program.

HELIO TOLLINI, Ph.D. North Carolina State 1969. Actual and optimal use of fertilizer.

Regression analyses relating corn yields to nitrogen and other inputs from survey and experimental data are used to explain the paradox of high marginal rates of return to fertilizer. Much of the gap can be explained in terms of the variance of net revenue increasing sharply with nitrogen levels, the analyst never knowing the true functional form, and the small share of total costs involved. Evidence that farmers with large corn enterprises are closer to having optimum sales and nitrogen levels supports the hypothesis that larger farmers can afford to acquire more information.

EDWIN TONNESON, Ph.D. Syracuse 1969. Public policy in competitive recovery of fish resources.

ROBERT J. TOWNSLEY, Ph.D. Iowa State 1969. A quadratic programming model for economic analysis of swine rations.

JOSEPH W. UHL, Ph.D. Michigan State 1969. Market organization, vertical coordination, and farm price stability in the fresh fruit and vegetable industry.

LAWRENCE VAN MEIR, Ph.D. Kansas 1969. A study of policy considerations in managing the Georges Bank haddock fishery.

HERMAN J. VAN WERSCH, Ph.D. Minnesota 1969. Land tenure, land use, and agricultural development, a comparative analysis of Messinia (Greece) and the Cape Bon (Tunisia).

PAUL D. VELDE, Ph.D. Purdue 1970. The impact of transit and nontransit transportation rate structures on the soybean processing industry.

This study examined the effects of removing transit rate privileges and locational processing capacity constraints on particular segments of the soybean industry. The linear programming model employed indicated that significant savings to soybean producers and consumers under transit arose from partially bypassing regional processing capacity restrictions. The total effect of eliminating all locational processing capacity restraints led to further benefits. Gains external to the model appeared to enhance the general advantages of transit.

LOWELL D. WOOD, Ph.D. California (Berkeley) 1969. An economic analysis of the planning and evaluation procedures employed by the United States Army Corps of Engineers with particular reference to the proposed Dos Rios project in Northern California.

The Dos Rios water development proposal has been attacked by economists for being economically unjustifiable, by conservationists for its detrimental environmental consequences, by Indians for proposing to inundate their land and homes, and by concerned people who question its actual need. The thesis is concerned with an economic analysis of these issues and on the Corps planning and evaluation procedures. The benefit cost technique is used as a framework for the analysis. The study is concerned with a broad, rather than a narrow point of view.

JUAN A. ZAPATA, Ph.D. Chicago 1970. The economics of pump irrigation: the case of Mendoza, Argentina.

The withdrawal of ground water from a common pool by several users, with no property rights defined over the water in the aquifer, can be expected to result in the well-known misallocative effects of common property resource exploitation. In this study a model is developed to estimate the externalities associated with the lowering of the water table caused by excess pumping. Then these externalities are estimated for an agricultural area in Argentina.

JOHN G. ZURENKO, Ph.D. Montana State 1969. Alfalfa dehydration: operations and costs.

Manpower, Labor, and Population; including Trade Unions and Collective Bargaining

ABDALLAH ABOU-AISH, Ph.D. California (Los Angeles) 1969. The international labor organization and management development in underdeveloped countries.

W. PAUL ALBRIGHT, Ph.D. State University of New York (Buffalo) 1970. The impact of the unions on wages of supermarket cashiers in southern Ontario.

TERRY M. ALDRICH, Ph.D. Texas (Austin) 1969. Rates of return earned on investment in technical education in the antebellum American economy.

This dissertation is an attempt to measure the rates of return earned on investment in formal technical education in the antebellum American economy. To accomplish the task, samples of West Point graduates (who were shown to be representative of all recipients of formal technical training) from the classes of 1817–35 were used. The costs of educating the cadets, and the increment to their productivity (as measured by the addition to their lifetime earnings) attributable to education were then estimated. The study concluded that returns to technical education were relatively low, and hence that it was a bad investment, contributing relatively little to early American economic development.

JEREMIAH M. ALLEN III, Ph.D. Colorado 1970. The return to human migration: a case study of a mass layoff.

RAMON C. ALONSO, Ph.D. State University of New York (Buffalo) 1970. A study of commitment orientations among professional personnel: nurses' commitment to the profession, clinical specialty, and employing organization.

BERNARD E. ANDERSON, Ph.D. Pennsylvania 1970. The Negro in the public utility industries: an investigation into the development and implementation of racial employment policy.

The study describes and analyzes the employment of Negroes in the investor-owned electric power, gas distribution, and telephone industries. An attempt is made to distinguish between the factors responsible for the development of employment policy regarding black workers, and changes in the level of employment of black workers. The separate identification of the two sets of factors allows one to better explain the variation in Negro employment among the three industries, and among firms in the same industry.

THOMAS ARNOLD, Ph.D. Syracuse 1970. The teamsters union as a determinant of the structure of the trucking industry.

DAN M. BECHTER, Ph.D. Yale 1970. The indirections of market forces as revealed in academic salary determination.

HERMAN BERLINER, Ph.D. City University of New York 1970. Real economic benefits of higher education.

BRUCE R. BOALS, Ph.D. Florida 1969. Some economic aspects of workmen's compensation in Florida.

The objective of this inquiry was to examine the efficacy and adequacy of a typical state workmen's compensation program. The treatment first provides a neoclassical analysis of the cost burden of work injuries. Next, the historical evolution of the program is developed. Appropriate provisions of the statute and the machinery for administration of the law are then analyzed and evaluated. Conclusion: while the program has been a limited success, many challenges for improvement remain.

JAMES J. BRADY, Ph.D. Notre Dame 1969. A case study of evolving standards in arbitration cases concerning layoffs.

THOMAS A. BREWER, Ph.D. Cornell 1969. Centralization of union-management bargaining and its relationship to performance in the Chicago, Detroit, and Buffalo fluid milk markets.

Highly centralized collective bargaining in the fluid milk industry provides the opportunity for firm owners to increase profits and for owners to increase or maintain employment and increase wage benefit levels. The higher degree of centralized collective bargaining in Chicago than in Detroit is reflected in higher unit labor costs in processing and wholesale distribution of fluid milk in the former city and in wider margins per quart in Chicago than in Detroit.

NEIL BUCKLEW, Ph.D. Wisconsin 1970. Effects of unionization of non-academic employees on university personnel administration.

PAUL L. BURGESS, Ph.D. Colorado 1969. Job vacancies in the Denver labor market.

DONALD R. BURKE, Ph.D. Pennsylvania 1970. The impact of unions on social welfare agency management.

This study ascertains the impact of social welfare unions on agency management in both the public and private sectors. Through an analysis of collective bargaining agreements and personal interviews, unions in the private social welfare agencies were found to be particularly effective when compared with unions in education and private industry. Unions in public agencies also compared favorably to teacher unions but have not been as effective as unions in private agencies and private industry.

JAMES H. H. CARRINGTON, Ph.D. American 1970. A critical analysis of the concept of representation under the National Labor Relations Act, as amended.

The author first constructs a model of the concept of representative government. He then identifies five institutional forces acting upon representative government in labor unions. The effects on representation of each of the five forces is analyzed and compared with the model. Finally, the concept, as it has been interpreted, is examined from the viewpoint of the public interest.

GENE L. CHAPIN, Ph.D. Duke 1969. A study of labor market information theory in the South African economy.

PETER Y. COMAY, Ph.D. Princeton 1969. International migration of professional manpower: the Canada-U.S. case.

The objective of this study was to assess the importance of various influences on the migration of professional manpower between Canada and the United States. The human capital model was used to derive hypotheses about the migrational decision of such individuals; these were then tested with survey data. Results were presented separately for the decisions to ever migrate, to migrate in a given year, and to return. Our findings were then used to quantify the benefits and costs of Canadian emigration and Canadian study abroad.

PHILIP G. COTTERILL, Ph.D. Northwestern 1969. A model of labor in retail trade.

A model of the supply and demand for labor was developed and estimated for twelve types of retail trade establishments using SMSA data from the Census of Business. The estimated demand functions were used to infer expected employment changes resulting from representative increases in the minimum wage and the rate of growth in retail sales.

WILLIAM Y. DAVIS, JR., Ph.D. Georgia 1970. Labor force responsiveness to seasonal variations in employment opportunity.

This study offers a partial explanation for the large seasonal swing in labor force participation. Regression analysis tests the seasonal responsiveness hypothesis that a substantial number of workers enter and leave the labor force in response to seasonal changes in the demand for labor. A significant positive finding indicates that the labor force is highly responsive to seasonal changes in employment opportunity and also suggests the possible presence of seasonally discouraged workers.

EVANGELOS S. DJIMOPOULOS, Ph.D. Columbia 1969. Temporary labor migration: effects on the countries of emigration.

JOSEPH W. DUNCAN, Ph.D. Ohio State 1970. Technical manpower in a research organization: a case study of patterns of job preparation and job function.

Using both tabular and multiple regression analysis,

this study evaluates the relationship between the specific job functions of scientific and technical manpower and their patterns of educational attainment and work experience. The analysis is based upon mail questionnaire responses from 1154 male scientists, engineers, and technicians (a response rate of 56 percent) employed by a large contract research organization. A model was developed which viewed education, experience, and ability as substitutes in the knowledge production function, the "product" of the research industry.

B. CURTIS EATON, Ph.D. Colorado 1969. Reemployment as a probability process: a case study of a mass layoff.

This study is an analysis of the reemployment experience of 5,000 aerospace workers. The reemployment process is investigated as a probability process. Three different models are developed. The main conclusion of interest is that random factors play a major, perhaps predominant, role in the process.

OLANREWAJU J. FAPOHUNDA, Ph.D. New York 1970. Manpower planning in Nigeria: a case study of the means of solving unemployment problems in developing countries.

The study examined the chief means of acquiring skills in Nigeria—primary education and the apprentice system—and evaluated the effectiveness of manpower planning in easing unemployment. Unemployment persists because the prerequisites for a successful manpower planning—wage employment, development programs, good educational system, a manpower planning body—are poorly coordinated and developed. A model developed to measure the burden of free primary education showed that Nigeria is underinvesting in primary education and in tangible capital. Improvements are possible only if government policies are changed.

ERIC G. FLAMHOLTZ, Ph.D. Michigan 1969. The theory and measurement of an individual's value to an organization.

MARTIN J. GANNON, Ph.D. Columbia 1969. Productivity and employee turnover in a branch-banking system.

ALGIE R. GRIMES, Ph.D. Oklahoma. An analysis of the determinants of the commuting and residence patterns of the Oklahoma City Air Material Area labor force.

The objective of this study was to analyze the most strategic factors determining the drawing power of a large employer from a labor market area. The firm studied was the Oklahoma City Air Material Area (OCAMA). Factors influencing OCAMA's drawing power and the associated commuting and residence patterns of the OCAMA labor force were analyzed in three categories: characteristics of the OCAMA laborshed area; characteristics of OCAMA; and personal attributes of the OCAMA work force.

SAMUEL GUBINS, Ph.D. Johns Hopkins 1970. The impact of age and education on the effectiveness of training: a benefit-cost analysis.

An examination of institutional training MDTA to determine the impact of a changing clientele on the efficiency of training and the effects of age and previous education on benefits and costs of training. The analysis reveals continuing payoffs from training, despite the change in clientele from the technologically displaced to the hard-core unemployed, and significantly higher payoffs from training the less educated. Speculation on the impact of expanding institutional training on benefits and costs.

ALAN L. GUSTMAN, Ph.D. Michigan 1969. Human capital, capital theory, and occupational earnings.

MARY T. HAMILTON, Ph.D. Pennsylvania 1969. A study of wage discrimination by sex: a sample survey in the Chicago area.

The study focuses on the estimation of "pure" measures of wage discrimination on the basis of sex, within narrowly defined occupations. Using multiple regression techniques, other factors such as age and education to which wage differentials might be attributed are taken into account. The results of the analysis clearly suggest that wage discrimination has a sex dimension. In addition, the estimated sex-wage differentials generally exceed those attributable to color, and often by considerable amounts.

WILLIAM G. HAMM, Ph.D. Michigan 1969. Race, school quality, and achievement in northern United States.

HERBERT HENEMAN, Ph.D. Wisconsin 1970. An empirical investigation of expectancy theory predictions of job performance.

JAMES K. HIGHTOWER, Ph.D. Claremont School 1970. A Markov chain analysis of labor market activity.

The Theil-Van De Panne quadratic programming algorithm was employed to derive a computational scheme for obtaining restricted least-squares estimates of Markov chain transition probabilities from time-series data. This technique was then applied to monthly data to estimate short-run transitional behavior of the civilian employable population with respect to labor market activity. Results were interpreted in terms of some current extensions of the Woytinsky-Humphrey debate.

WALTER HUBNER, Ph.D. Wisconsin 1969. Individual need satisfaction in work and non-work: a comparative study of the effects of the technology organization of work.

JOHN L. IACOBELLI, Ph.D. Texas (Austin) 1969. Training programs of private industry in the greater Cleveland area.

This is a study of current policies, practices, and attitudes of employers in training normal and disad-

vantaged labor. Employers reveal what they consider as proper manpower roles for government, private industry, and the educational system. Subjects investigated are: skill shortages, training by private industry versus other labor procurement methods, factors influencing training, industry's manpower projections, problems with disadvantaged labor, impact of government training programs, and federal financial incentives for training. Included are national manpower policy recommendations to achieve better matching of future supply with demand for trained labor and to improve the coordination of private and government training programs.

NEMATALLAH N. IBRAHIM, Ph.D. Minnesota 1970. A comparison of the economic returns to regular and vocational high school education.

This thesis evaluates several educational-occupational "tracks." Each track represents one type of high school education (regular academic, regular with substantial non-specialized vocational training, and specialized vocational) followed by a given type of occupation (vocational or non-vocational). Wage differentials between any two tracks (over a 5-year period) and public per student costs were estimated. Net benefit differentials for and between tracks were calculated and used to evaluate the present mix of high school education from both individual and social viewpoints.

BARBARA JANOWITZ, Ph.D. Johns Hopkins 1970. An empirical study of the effects of socioeconomic development on fertility rates.

This thesis examines the validity of employing results obtained from cross-section studies to predict the effects of changing socioeconomic conditions on fertility rates. In order for cross-section studies to be able to substitute for time-series studies, it must be assumed that both cultural and historical conditions play no role in affecting fertility rates. However, the results indicate that the effects of socioeconomic development on fertility rates cannot be considered in a cultural vacuum, that cultural conditions do affect fertility rates.

PHILIP W. JEFFRESS, Ph.D. Kentucky 1969. Racial employment patterns in the urban mass transportation industry.

This dissertation is concerned with the problem of identification and analyses of the factors affecting racial employment patterns in the urban mass transit industry. The study is based on a survey of selected firms in the industry, labor unions, equal employment opportunity agencies, and other civil rights groups. Subsequent to a historical treatment of social employment patterns, the study concentrates on the major determinants of these patterns. These include manpower and recruitment policy, labor demand, industrial location, union policy, and fair employment legislation.

MATHEW W. JEWETT, Ph.D. Alabama 1969. An analysis of arbitration in Southern Bell Telephone and Telegraph Company.

THOMAS JOHNSON, Ph.D. North Carolina State 1969. Returns from investment in schooling and on-the-job training.

Assuming a continuous investment in human capital function, non-linear regression techniques are used to predict lifetime earnings. With this framework investigation of the relation of schooling and on-the-job training with regard to race and region are made.

JOHN L. KNAPP, Ph.D. Virginia 1970. Wages in Virginia relative to the nation, 1939 to 1965.

The study measures and explains reasons for changes in wages in Virginia relative to the nation from 1939 to 1965. After a description of measurement problems and a review of the literature, eleven factors are examined in order to determine their gross and net effects on relative wages. A composite index including most major employment sectors showed little change in Virginia wages relative to the nation from 1939 to 1965.

ROSLYN KUNIN, Ph.D. British Columbia 1970. Labor force participation rates and poverty in Canadian metropolitan areas.

A model is developed explaining both discouraged and added worker behavior. Regression analysis is applied to census tract data for major Canadian cities to discover the determinants of labor force participation. Separate regressions are run for high, middle and low income tracts. Unemployment, wages and sociodemographic variables are important in determining participation rates. Their influence differs significantly among income groups. A strong discouraged worker effect is found.

JOSEPH SHING LEE, Ph.D. Massachusetts 1970. An econometric theory of union growth.

ROBERT I. LERMAN, Ph.D. Massachusetts Institute of Technology 1969. An analysis of youth labor force, school activity, and employment rates.

This dissertation examines a number of hypotheses concerning youth labor force participation, school activity, and employment with the use of data from the March 1967 Current Population Survey. Regressions across individuals include as explanatory variables individual, family, and SMSA characteristics. Examples of results are discouraged and added worker effects by race and by school activity, and the influence on youth employment of the family head's occupation.

SAMUEL LEVY, Ph.D. Wayne State 1969. Some aspects of the international migration of human capital: the case of British physicians.

JOHN M. LEYES, Ph.D. Iowa State 1969. The market for public school teachers in Iowa: an economic analysis.

The Iowa teacher market was evaluated through the use of primary data obtained in a random sample of 59 school districts that included a survey of school administrators and of school teachers with zero years of tenure. The objectives included: description of the market participants; reasons for teacher resignations; importance of teacher choice or decision variables; search channels and search costs; teacher misemployment; employment opportunities competitive with teaching; teacher turnover; and salary as a market variable.

THEODORE P. LIANOS, Ph.D. North Carolina State 1969. Greek labor migration to Western Europe.

Two problems are investigated. One, the effect of pecuniary to nonpecuniary factors on migration and the implied psychic cost of moving is examined by a regression of moves on both kinds of variables. Second, the migration process is looked at as a stock-flow problem, and a distributed lag in the process is analyzed.

DONALD MARTIN, Ph.D. California (Los Angeles) Claims to work opportunity: an economic analysis of alternative configurations of hiring rights.

JOHN P. MATTILA, Ph.D. Wisconsin (Madison) 1970. Theory and estimation of quit functions.

The thesis develops a theory of the decision to quit a job emphasizing the search process and job opportunity. Theoretical and empirical evidence suggests that the majority of quitters move directly to new jobs without intervening unemployment. This has important implications for mobility, turnover costs, and wage determination (and the Phillips' Curve). A regression analysis of B.L.S. quit rate data (1958–67) by 2-digit industry provides further insights into quit behavior.

EDYTHE S. MILLER, Ph.D. Colorado 1970. The concept and measurement of poverty in the United States.

RAUL MONCARZ, Ph.D. Florida State 1969. Effects of environmental change on human capital among selected skilled Cubans.

Using regression techniques, this thesis traces the effects of various kinds of labor market restrictions on the incomes of skilled Cubans. These include licensing, fixed educational requirements in the U.S., state registry, and residency requirements. In general, it was found that where the entry difficulties exist, especially in licensing, the adjustment losses on income were greatest. Residency, accreditation, and educational requirements were also significant barriers. The extent to which the provision of education, training, market information, placement facilities, and mobility affect income was also explained.

MICHAEL L. MOORE, Ph.D. Michigan 1970. Managerial learnings and socialization.

WILLIAM J. MOORE, Ph.D. Texas (Austin) 1970. The growth and development of teacher unions in the public schools: a theoretical interpretation.

This study attempts to identify the major factors leading to the origin and development of unions among public school teachers in the United States. The study sets forth a quantitative history of teacher organizations and presents a conceptual framework for analyzing the various factors influencing the teacher's decision to join a union. It also traces the history of the teacher union movement in detail and examines the factors which influence union growth.

BETH NIEMI, Ph.D. Columbia 1970. Sex differentials in unemployment in the United States and Canada, 1947–66.

This study attempts to explain the fact that women have had a higher unemployment rate than men in the United States throughout the postwar period. Three factors—a high rate of inter-labor force mobility, a relative lack of specific job training, and a tendency to be geographically and occupationally immobile—are postulated as being responsible for the high unemployment rate of women relative to men. These hypotheses are tested empirically, and the results thus obtained are also utilized in the analysis of data on two subjects closely related to the main topic of this study: the structure of differential unemployment in Canada, and the changes that have occurred during the postwar period in differential unemployment in the United States.

HAROLD OAKLANDER, Ph.D. Columbia 1969. Some unanticipated effects of advanced education on a critical manpower resource: the public school teacher.

CHARLES F. O'DONNELL, Ph.D. Fordham 1970. Old-age and survivors insurance benefits in a macrodynamic context.

PAUL P. N. OFFNER, Ph.D. Princeton 1970. Labor force participation in the ghetto: a study of New York City poverty areas.

This dissertation examines the relationship between labor supply and location in urban poverty areas. The empirical findings are based on cross-sectional regressions employing 1960 New York City census tract data. Distance to major centers of employment and the availability of jobs in the immediate neighborhood are found to have a significant impact on labor supply. The results suggest that residential segregation makes a significant contribution to the employment problems of the Negro urban poor.

TAI OH, Ph.D. Wisconsin 1969. The role of international education in the Asian brain drain.

DANIEL A. ONDRACK, Ph.D. Michigan 1970. An investigation into changes in the level of authoritarianism among college student populations from 1958 to 1968.

JUNE C. O'NEILL, Ph.D. Columbia 1970. The effect of income and education on interregional migration.

KEITH S. PANTON, Ph.D. Washington State 1970. The development and impact of collective bargaining in the bauxite and alumina industry in Jamaica, 1952–68.

JOHN H. PENCAVEL, Ph.D. Princeton 1969. An analysis of the quit rate in American manufacturing industry.

In this study into voluntary labor turnover, the author draws upon recent work in the theory of human capital to derive hypotheses concerning interindustry quit rate behavior. These hypotheses are then tested with data from the Census of Population and Bureau of Labor Statistics. Other hypotheses and possible objections to the econometric techniques are considered. Finally, a tentative explanation is offered for the negative postwar trend in the quit rate.

NICHOLAS S. PERNA, Ph.D. Massachusetts Institute of Technology 1969. Wage determination in a nonunion firm.

DEWAYNE J. PIEHL, Ph.D. Michigan 1970. Some organizational constraints on the business decision making process.

LAWRENCE D. POSNER, Ph.D. Harvard 1970. Analysis of investment in computer programmer training.

FREDERICK A. RAFFA, Ph.D. Florida State 1969. An analysis of the determinants of labor migration in the state of Florida, 1961–65.

Using a regression model, this thesis attempts to analyze those factors which affect migration to and within Florida—a high internal and external migration state, isolating the effects for major counties of such variables as levels of industrialization, service activity, levels of employment, unemployment rates.

DALE RASMUSSEN, Ph.D. Southern Methodist 1969. Determinants of rates of return to investment in on-the-job training.

This study evaluates public investment in on-the-job training (OJT) and tries to identify systematic determinants of the "profitability" of this investment. Information for a national sample of trainees completing OJT was obtained from the U.S. Department of Labor and additional data was selected from the 1-in-1,000 1960 Census sample and are then adjusted to the regional level using wage survey indices.

DAVID W. RASMUSSEN, Ph.D. Washington (St. Louis) 1969. The determinants of the nonwhite/white income ratio.

The relative income of nonwhite males is investigated via regression analysis using cross-section data for states, SMSAs and occupations for 1960. The trend in the nonwhite/white income ratio from 1948 to 1964 is also investigated. Estimates of the impact of discrimination on the economic well being of nonwhites emerge from the analysis.

V. LANE RAWLINS, Ph.D. California (Berkeley) 1969. Government sponsored training programs for the disadvantaged youth as a part of efficient long-run manpower policy.

BERNARD D. ROSTKER, Ph.D. Syracuse 1970. Manpower theory and policy and the residual occupational elasticity of substitution.

BRIAN S. RUNGELING, Ph.D. Kentucky 1969. Impact of the Mexican alien commuter on the apparel industry of El Paso, Texas.

Recent criticism of Mexican commuters results from their impact on U.S. border labor markets. This paper analyzes these effects as well as establishing that commuters are an important part of the area's work force. Findings of the study showed: commuters depress wages as they are part of the large pool of unskilled workers; commuters have higher real wages than the U.S. workers in the same industry. Solutions to the large labor supply center on job creation, restriction of entrance, and improvement in the operation of the labor market.

LARRY B. SAWERS, Ph.D. Michigan 1969. The labor force participation of the urban poor.

JOSEPH SCHACHTER, Ph.D. City University of New York 1969. Capital values and relative wage effects of immigration into the United States, 1870–1930.

LOREN C. SCOTT, Ph.D. Oklahoma State. An economic evaluation of on-the-job training conducted under the auspices of the Bureau of Indian Affairs in Oklahoma.

This study evaluated the economic effectiveness of on-the-job training received by 226 American Indians between 1960 and 1967 under the auspices of the Bureau of Indian Affairs. Net private and social returns are estimated. Several social benefit-cost ratios were calculated using different combinations of discount rates and time horizons. The most conservative ratio computed was 7.6, combining a 10 percent discount rate with a five-year time horizon. A ratio of 29.4 was derived using a 6 percent discount rate and a thirty-six year time horizon. Evidence indicates that the training periods negotiated by the BIA were unnecessarily long to accomplish program objectives.

JAMES R. SELDON, Ph.D. Duke 1969. Some aspects of Canadian immigration policy in theory and in practice.

GEORGE B. SIMMONS, Ph.D. California (Berkeley) 1969. The Indian investment in family planning.

JACK SNYDER, Ph.D. Wisconsin 1969. Personal and behavioral facts affecting the supply of the self-employed.

PAUL D. STAUDOHAR, Ph.D. Southern California 1969. Collective bargaining, disputes, and dispute settlement procedures in public employment.

This study analyzed certain collective bargaining and dispute settlement procedures currently in existence in jurisdiction at state and local levels of government. The objective was to determine which procedures might be included in or excluded from viable systems of collective bargaining in public employment. Voluntary binding arbitration appeared to be the most efficacious means of settling disputes in public employment, but its use has been hampered by its dubious legality in many jurisdictions.

WILLIAM H. STEVENSON, Ph.D. Vanderbilt 1969. Shift patterns in Tennessee Valley region manufacturing employment; 1950–75.

MYRA H. STROBER, Ph.D. Massachusetts Institute of Technology 1969. Relationships between the level of economic development and the interindustry wage structure.

The major finding is that the industrial (two-digit manufacturing) earnings hierarchies of underdeveloped countries are poorly correlated with that of the United States; but, that as a country develops, its earning hierarchy becomes more like that of the United States. Particular stress is placed on the importance of productivity structures in explaining wage structures. The study uses U.N. and I.L.O. data on average earnings by industry.

ABDEL-RAHMAN TAHA, Ph.D. California (Los Angeles) 1970. Sudanese labor movement.

MARCEL TENENBAUM, Ph.D. Columbia 1969. A demographic analysis of inter-state labor growth rate differentials; United States 1890–1900 to 1940–50.

The intercensal growth rate of the labor stock of a state has been decomposed into the contributions of natural increase, migration and changing participation rates. This was done for each decade from 1890–1900 to 1940–50, by states and for many age-sex groups. Regression techniques were then used to analyze relations among these components of labor stock change; special attention was given to the effect of data estimation errors on the regression coefficients.

ARMAND J. THIEBLOT, Ph.D. Pennsylvania 1969. Policy formation and managerial reaction: Negro employment in the banking industry.

Although no standards exist to compare Negro employment participation levels in banking, the levels themselves are documented and analyzed. Factors found having greatest overall social postures and labor market conditions, institutional size, government pressures, and the willingness of banking officers to become involved. From fewer than 2 percent Negro employees in 1960, banking had almost 6 percent in 1968, and the rate of increase was impressive.

ROBERT THORNTON, Ph.D. Illinois (Urbana) 1970. Collective negotiations for teachers: historical and economic effects.

ROBERT D. TOLLISON, Ph.D. Virginia 1969. An analysis of the taxation and collective choice aspects of the military draft.

ANDREAS C. TSANTIS, Ph.D. Wisconsin (Madison) 1970. The internal and external Greek labor migration in the postwar years.

This dissertation studies the determinants of Greek interregional male and female migration in 1956–61 and of the external migration of Greek males and females to six recipient countries in 1955–66. The interaction of internal and external migration and the rural/urban origin of the emigrants are also investigated.

ARVIL VAN ADAMS, Ph.D. Kentucky 1970. A study of Negro employment patterns in metropolitan Memphis, Tennessee.

The objectives of this study were to describe by occupation and penetration the employment patterns of Negroes in metropolitan Memphis, Tennessee, and to identify and analyze factors explaining these patterns. The primary data source was employer data collected under Title VII of the 1964 Civil Rights Act. Included in the set of factors analyzed were education, migration patterns, housing segregation, job information, unions, economic growth, firm size, and capital intensity. Institutionalized forms of discrimination were found to be the major determinants of Negro employment patterns in Memphis in contrast to overt acts of employer discrimination.

WILLIAM E. VICKERY, Ph.D. Chicago 1969. The economics of the Negro migration 1900–60.

This study seeks to identify and measure the major determinants of rates of Negro interstate migration and its allocation during 1900–60. This is achieved by specification of a migration model which permits econometric estimation of elasticities of these rates and allocations with respect to various explanatory variables for each state in each intercensal decade, 1900–10 to 1950–60. Relative strengths of explanatory variables are determined by partitioning techniques used in many production function analyses.

KEITH VOELKER, Ph.D. Wisconsin 1969. The history of the international brotherhood of pulp, sulphite, and paper mill workers.

STANLEY S. WALLACK, Ph.D. Washington 1969. A disaggregated study of market and non-market wage changes in U.S. manufacturing.

The study presents an operational model to explain the market and non-market causes of the percentage rate of change of money wages at the industry level. Five variables were selected to represent market causes: gross new hire rate, percentage change between value-added per man-hour and market wage rate, difference between gross new hire and quit rates, lagged profit rates, and percentage change in the con-

sumer price index. The two non-market causes were the wage-price guidelines and wage spillover.

NEIL S. WEINER, Ph.D. Harvard 1970. An empirical analysis of the markets for construction material and labor inputs.

This thesis develops econometric estimates of both the supply and demand functions for major building materials and for regional contract construction labor markets. The nature of available data imposes somewhat hybrid functional specifications. For materials, annual observations suggest rather elastic supply conditions, except for the more concentrated industries. For construction labor, the monthly model appears to be appropriate for the Northeast and Northcentral regions, but not for the South and West, primarily due to the density of data coverage and regional differences of industrial structure.

JOSEPH P. YANEY, Ph.D. Michigan 1969. Organizational support and employee performance.

ALBERT ZUCKER, Ph.D. Columbia 1970. Some aspects of the economic effects of minimum wage legislation: 1947–1966.

Cross-section and time-series regressions are performed on national and state industries to ascertain the employment effects of the Fair Labor Standards Act. Wage elasticities of employment response are estimated. Intra-labor substitution is investigated to determine elasticities of substitution between labor factors. Data by sex and occupation is utilized.

LAWRENCE S. ZUDAK, Ph.D. Purdue 1969. A reinterpretation of the theory of labor markets with reference to the Midwest steel industries.

Marginal theory proves inapplicable to a steel mill in the contract period. Wages are fixed. Related jobs are hierarchically organized into promotional sequences so each job trains workers for the next. Skilled labor is supplied by vertical mobility based on sequence service. Unskilled workers are allocated between competing sequences by their maximization of expected mill-time income. Labor demand is fixed at each output level (mix) by standard crew agreement. Labor market equilibrium is Keynesian, underemployed.

Welfare Programs; Consumer Economics

SUSAN R. ACKERMAN, Ph.D. Yale 1970. The demand for used automobiles in the United States: 1954–66.

WILLIAM T. ANDERS, JR., Ph.D. Alabama 1969. The guaranteed minimum income: an analysis of its conceptual and economic origins.

MARVIN R. BURT, D.B.A. George Washington 1969. A critical appraisal of program analysis in the federal government: with particular attention to health programs.

The dissertation seeks to answer the research question: "What methodologies and approaches are and should be used in conducting health program analyses?" Included is a critical evaluation of the current "state-of-the-art" in medical program analysis in the federal government with specific constructive suggestions for improvements.

ROBERT F. COOK, Ph.D. Lehigh 1970. Retirement and stabilization: an economic analysis of the retirement test.

The study is divided into two parts. The first is to estimate the effects of private pensions, changes in labor demand, and changes in legislation on rates of retirement under the Railroad Retirement and OASI programs. The second is concerned with whether the benefit structures, through the earnings test, operate to stabilize income from wages and benefits. The same analysis is then applied to the taxing structures and the programs as a whole.

WILLIAM C. DUNKELBERG, Ph.D. Michigan 1969. Forecasting consumer expenditures with measures of attitudes and expectations.

ROBERT G. EVANS, Ph.D. Harvard 1970. Efficiency incentives for hospital reimbursement.

Rapidly rising hospital spending appears to be related to increasing inefficiency rather than increasing output from the hospital industry. Reimbursement incentives seek to make the industry's access to economic resources conditional upon improved efficiency rather than simply supplying whatever the industry asserts its costs to be. Such incentives depend on a realistic specification of the structure and goals of the hospital management group and of the technical constraints under which they operate. This thesis attempts to specify the objectives of hospital management and then evaluates proposed incentive plans by predicting managerial responses. Some empirical analysis of technical constraints is presented but this is suspect because the measured behavior of an industry which may not be minimizing its costs does not necessarily reflect technical constraints. Finally suggestions and guidelines are provided for future experimentation.

GILBERT R. GHEZ Ph.D. Columbia 1970. A theory of life-cycle consumption.

A theory of life-cycle consumption is developed which carries markedly different implications than the standard Fisher-Modigliani-Brumberg model. Households are assumed to engage in nonmarket activities, combining the services of market goods and their own time. The model explains why the shape of the life-cycle consumption path is not independent of the shape of the wage rate profile, even in the absence of unexpected changes in income. The strong implications of the model were found to be consistent with observed expenditure patterns in 1960, using the BLS Survey of Consumer Expenditures.

MICHAEL GROSSMAN, Ph.D. Columbia 1970. The de-

mand for health: a theoretical and empirical investigation.

The aims of this study are to construct and estimate a model of the demand for the commodity "good health." It is assumed individuals inherit an initial stock of health that depreciates over time and can be increased by investment. Age via its effect on the rate of depreciation on health capital, and education via its effect on the efficiency of the health production process, are shown to influence the shadow of price of health. The empirical sections of the study estimate demand curves for health and medical care and health production functions.

SAMUEL B. HINCHEY, Ph.D. Missouri (Columbia) 1970. An analysis of interregional variation in general hospital expenses.

Interregional variation in total expense per patient day for non-federal, short-term general hospitals in the continental United States was analyzed using a multiple regression, multiple equation model and three-stage least squares estimation technique. The estimated model supports the hypothesis that strong relationships exist between hospital costs and personal income, availability of physicians, average plant scale and the nursing wage rate. Availability of hospitalization insurance, the retail price level, and federal expenditures for hospital construction proved to bear negligible relationships to cost levels.

BARBARA H. KEHRER, Ph.D. Yale 1970. The nursing shortage and public policy: an economic analysis of the demand for hospital nurses in Connecticut.

GERALDINE P. KNIGHT, Ph.D. Southern California 1970. An empirical study of some significant variables in consumer expenditures: Los Angeles County 1970.

One thousand personal interviews serve as the basis for a micro-analysis about allocations of scarce resources by the Los Angeles County consumer in February 1970. Using a stratified random probability sample, the questionnaire provides data revealing the range and depth of the effect from relevant demographic and environmental factors influential in consumer expenditures for durable and non-durable goods and services. The results are salient in an inflationary economy with increasing personal income.

STANLEY LONG, Ph.D. Iowa 1969. Demand for medical care.

IRENE LURIE, Ph.D. California (Berkeley) 1969. An economic evaluation of aid to families with dependent children.

JOHN A. RAFFERTY, Ph.D. New York 1970. Hospital use patterns and measurement in terms of case-mix.

A welfare theory framework is used to predict case-mix variations in both short-run (monthly) and long-run situations (increased capacity). Hospital use by diagnosis is examined, confirming the expected utilization patterns. A case-mix index is consequently developed, and its uses are illustrated in empirical applications. Implications for future costs studies are suggested.

UWE E. REINHARDT, Ph.D. Yale 1970. An economic analysis of physicians' practices.

RUDOLPH T. ROBERTS, Ph.D. Nebraska 1969. Public assistance and work effort: a quantitative cross-sectional analysis.

SHYAMALENDU SARKAR, Ph.D. Michigan State 1969. The copper country medical industry of Michigan as it serves rural people.

JUDITH A. SEGAL, Ph.D. George Washington 1970. The allocation of public funds to improve the diet of low-income families.

HARRY L. SHUFORD, Ph.D. Yale 1970. Subjective variables in economic analysis: a study of consumers' expectations.

BRUCE C. STUART, Ph.D. Washington State 1970. Distributive equity in the financing of American medical care, 1950 and 1960–61.

HOWARD P. TUCKMAN, Ph.D. Wisconsin (Madison) 1970. A study of college choice, college location, and future earnings.

The first portion of this study examines demand for college in areas containing a college as compared to demand in other areas. The second develops two theories of the determinants of college choice. One, a consumption theory, assumes students choose a college for the satisfactions received while at college. The second, an investment theory, assumes students choose colleges on the basis of expected future returns. Two models are developed and tested using Wisconsin data.

KENNETH L. WERTZ, Ph.D. Carnegie-Mellon 1970. Four essays upon durable goods, second-hand markets, economic intermediaries, and the pricing of undergraduate textbooks.

JACK W. WHITE, D.B.A. George Washington 1969. Administrative problems encountered in the operation of community mental health centers.

The study identifies administrative problems from two perspectives, those which are common to a majority of the participating community mental health centers, and those which are disruptive to the service functions of centers. The study also seeks to determine the extent to which disruptive problems vary in regard to the following selected characteristics of centers: administrative structure, ownership, demographic location, geographic location, catchment area, and policy-making body.

ROBERT C. YOUNG, Ph.D. Indiana 1969. Income and health insurance.

A study of the determinants of benefits paid out under Indiana University's health insurance program (as a case study of a relatively "comprehensive" Blue-Cross-Blue Shield Program) along with a study of the implications of this study for policy purposes.

MICHAEL ZUBKOFF, Ph.D. Columbia 1969. Increasing hospital efficiency through third party reimbursement plans.

Urban and Regional Economics

RICHARD J. AGNELLO, Ph.D. Johns Hopkins 1970. The second Chesapeake Bay Bridge: a cost-benefit study in public highway investment.

Using both static and dynamic user benefit efficiency criteria, the dissertation measures the cost and benefits of a parallel Chesapeake Bay bridge. Benefits are calculated by constructing data on time losses in traffic queues on the existing bridge, and estimating traffic demand functions using multivariate linear regression. Issues in highway studies are discussed, and estimates for value of travel time, optimal short-run efficiency tolls, and self-financing tolls are calculated.

LEONARD D. ATENCIO, Ph.D. Kansas State 1969. Forecasting from a regional input-output model: Kansas 1975.

This study designs and implements a forecasting system, applying the Leontief Open Input-Output Model to the state of Kansas. The final demand vector contains separate estimates of the individual components. The household and governmental sectors are estimated by regression techniques. The export sector is forecast by a matrix of direct export requirements, using the input-output forecasting techniques. The matrix of direct input requirements uses the "best practice" firms within each industry as the average input coefficient for the forecast year. The forecast final demand and direct requirement matrix are combined as the input-output system to generate forecasts from the base year (1965) to the forecast year (1975).

CLYDE T. BATES, Ph.D. Kentucky 1969. The effect of a large reservoir on local government revenue and expenditure.

This study investigates the effect of a large multipurpose reservoir on the revenue and expenditure of county governments and school districts that lost significant amounts of their tax base from the tax rolls. A case study was made of the three largest reservoirs constructed in Kentucky from 1957 to 1968. Time trends in tax severity were used to determine tax effects on local governments. These time trends evidenced no increase in tax severity or diminished expenditure resulting from construction of the reservoirs.

PAUL BAUM, Ph.D. California (Los Angeles) 1969. The relationship between city size and welfare.

BURLEY V. BECHDOLT, JR., Ph.D. Southern California 1970. Some external diseconomies of urban growth and crowding: Los Angeles.

The purpose of this dissertation is to summarize economic theoretical concepts related to external effects of urban growth and crowding, and relate them to Los Angeles. It is demonstrated that crime rates are highest where there is a high rate of male unemployment, high population density, overcrowding, and a low percentage of upper income families. Crime rates appear to be related more strongly to unemployment than to poverty, indicating that crime might be reduced through effective employment programs.

GEORGE A. BENZ, Ph.D. Oklahoma 1969. A "single tax" as a means of support for a local government: Edmond, Oklahoma.

This study determined the site value in the form of Henry George's unearned increment in Edmond, Oklahoma from 1941 to 1965. A tax if placed upon the site value would yield a return of two to three times greater than local expenditures. The conclusion drawn was that there was sufficient unearned increment in the form of land rent to support the local government, with additional monies left over to help support political subdivisions.

MICHEL BEUTHE, Ph.D. Northwestern 1968. Freight transportation mode choice: an application to corn transportation.

The study proposes a model of choice of a transportation mode for freight in perfect competition, which takes into account all transportation costs including time and quality differences, and uses the analysis of discrimination to forecast choices. As a by-product it develops estimates of quality differences. Second, the study uses and develops spatial analysis to forecast choices by regions.

WILLARD R. BISHOP, JR., Ph.D. Cornell 1969. An application of the intra-urban gravity model to store location research.

The primary objective was to test the validity of the original and two modifications of the intra-urban gravity model as a sales estimation technique for supermarkets. The models were tested in a large metropolitan area consisting of eighty supermarkets. All three forms of the model were unsatisfactory. Sales estimates produced by the models ranged from 24 to 231 percent of actual sales.

SAM H. BOOK, Ph.D. Columbia 1970. Costs of commuters to the central city as a basis for commuter taxation.

The author attempts to develop, verbally and diagrammatically, a model explaining the effects of the existence of a given commuting population on the resource costs of providing urban services in the central city servicing those commuters. From the model an approach is suggested for determining central-city service costs attributable to commuters. A disaggregative

empirical analysis of selected services in New York City is then performed as an example of this approach.

RICHARD G. BRUINS, Ph.D. Washington (St. Louis) 1970. Federal aerospace procurement and local income: a case study of the aerospace industry in St. Louis.

LOUIS P. CAIN, Ph.D. Northwestern 1969. The sanitary district of Chicago: a case study of water use and conservation.

William Dean's geographic approach to economic location theory was used to analyze Chicago's site, and it is argued that the site is economically defensible on theoretic terms. Given this, an investigation of Chicago's sequential decision-making with respect to water supply and waste disposal revealed the problems the city had to solve for the site to support economic growth. Since 1890, the sanitary district has been the agency responsible for significant decisions regarding these problems.

JAMES A. CHALMERS, Ph.D. Michigan 1969. A model of state and local government portfolio and real expenditures behavior: 1952–65.

RICHARD S. COWAN, JR., Ph.D. Lehigh 1970. Application of a modified employment shift model to forecast regional change.

This study modifies an employment shift model to forecast total work force employment in the Allentown-Bethlehem-Easton, SMSA to 1975, including the national economy and the changing internal structure of the region in the analysis. An index of homogeneity applied to these findings weights the strength of internal vs. external forces as they are expected to influence the growth process.

DENNIS COX, Ph.D. Stanford 1970. Economics of the distribution of population.

ROBERT T. CROW, Ph.D. Pennsylvania 1969. An econometric model of the northeast corridor of the United States.

The northeast corridor was defined to cover all states which are part of the U.S. "megalopolis" along the Atlantic Seaboard. A macroeconomic model of this region was constructed utilizing a comprehensive system of income and product accounts consistent with the national accounts. This model contains forty-three statistical equations and five identities. It was used for making annual forecasts, using an iterative technique to solve the system of non-linear equations.

RUSSELL L. DAWSON, Ph.D. California (Berkeley) 1970. The economics of regional planning for solid waste management.

ALLEN DOBSON, Ph.D. Washington (St. Louis) 1970. Price changes of single family dwelling units in racially changing neighborhoods.

The study used price indexes of single family dwelling units, estimated by a regression technique, which are adjusted for quality by sales-relatives, to test the hypothesis that house prices were affected by the process of integration in a suburban neighborhood in St. Louis County. None of the estimated indexes indicated that integrated homes sold at significantly different prices than did their nonintegrated counterparts. Hence the above hypothesis was tentatively rejected.

CHRISTOPHER M. DOUTY, Ph.D. Stanford 1970. The economics of localized disasters: an empirical analysis of the 1906 earthquake and fire in San Francisco.

This dissertation considers the effects of the exogenously caused, sudden destruction of economic resources within a single urban region. Included are investigations into the economic factors enabling San Francisco to survive; the determinants of the form and speed of the post-disaster recovery; and the recovery process itself, with emphasis on labor market phenomena. Relatively conventional economic theory and simple quantitative techniques applied to historical data are shown to be very useful in this context.

STEPHEN P. DRESCH, Ph.D. Yale 1970. Urban local government: aid, income, and patterns of fiscal activity.

DOUGLAS M. DUNN, Ph.D. Michigan 1970. Local area forecasting: an adaptive approach.

EDWARD G. EMERLING, Ph.D. Catholic 1970. Community organization for industrial development in selected New England states.

There are a variety of devices which are important to any well planned industrial development program. Among them are the Development Credit Corporation, the Industrial Building Authority, municipal bonding and property tax exemptions. These techniques are popular in most states but often have but limited use in New England. In the end, it is the people who are responsible for designing and executing the development programs who determine whether the effort will be a success or failure.

DONALD ERLENKOTTER, Ph.D. Stanford 1970. Preinvestment planning for capacity expansion: a multi-location dynamic model.

DONALD A. FREEDMAN, Ph.D. California (Los Angeles) 1969. The distribution lag as reflected by construction in the city of Los Angeles, 1955–1967.

DANIEL FREEMAN, Ph.D. Minnesota 1969. Interregional competition in producing, processing, and marketing snap beans.

DAVID GAYER, Ph.D. New York 1970. The suburban labor market: jobs for the low skilled.

The fact that many unemployed blacks are living in

central cities while numerous blue collar jobs are available in the suburbs has usually been attributed to the lack of public transportation and low-income housing in the suburbs. This thesis claims that a more important reason may be the relatively high costs imposed on poor job seekers in finding employment. These costs have splintered the metropolitan labor market into many isolated submarkets. To measure the interaction among submarkets and worker mobility an index of isolation was constructed for various classes of workers.

NORMAN J. GLICKMAN, Ph.D. Pennsylvania 1969. An econometric model of the Philadelphia area.

A twenty-six equation econometric forecasting model for the Philadelphia Standard Metropolitan Statistical Area is estimated by the method of two stage least squares. Endogenous variables include employment, output, personal income, gross regional product, population, and government expenditures and revenues. Following a discussion of the statistical problems involved in estimating small-sample models, short- and long-term forecasts and several public policy simulations are performed with the model.

JULIUS GYLYS, Ph.D. Wayne State 1969. Allocative problems of sheriff's police services in a metropolitan setting.

JOHN M. HARTWICK, Ph.D. Johns Hopkins 1969. Regional analysis by means of interregional input-output models and linear programming with applications to eastern Canada.

The static Isard interregional input-output model is reconstructed in terms of supply coefficients and technical coefficients and a detailed comparison is made of its relationship to the Chenery-Moses interregional input-output model. The Chenery-Moses model is utilized to generate regional multipliers by sector for eastern Canada. The concept of effective natural protection (protection arising from transportation costs) is developed in the context of locating firms competing in space.

RICHARD K. HAY, Ph.D. Kansas State 1970. The public works and economic development act of 1965: a theoretical analysis and evaluation.

A three-sector generalized evaluative model is used to determine 1) the theoretical consistency of the major provisions of the Act, 2) the tendency of these provisions to provide an "optimal" interregional allocation of resources and 3) the ability of social overhead infrastructure to induce regional growth. Certain inconsistencies within the Act are identified, and it is demonstrated that the ability of social infrastructure to induce regional growth is relatively weak, particularly in regions facing considerable resource deprivation.

DARYL A. HELLMAN, Ph.D. Rutgers 1970. A critical analysis of employment projection methods: a test case of New Jersey.

Projections of employment to 1975 for three-digit manufacturing industries are derived, using an explicit shift-share model. Using the localization quotient, industries are divided into import, export, and local market oriented. For the latter group, the state's shifting share of national employment is a function of shifting market share. For export industries, the share is related to indices of agglomeration economies in the state. The projections are evaluated, comparing performances among several models.

FREDERICK C. KIRBY, Ph.D. Wayne State 1969. A statistical analysis of urban growth characteristics.

THOMAS A. KLAASEN, Ph.D. Michigan State 1969. Alternative models of regional comparative advantage in the United States.

The Heckscher-Ohlin and classical trade models were tested empirically using U.S. regional data. A variable not usually associated with the two models, the "coefficient of resource dependency" was incorporated. It was concluded that an industrially developed region can be expected to display patterns of industry specialization as predicted by either model. For a developing region, initial attraction of industries is likely to be based on sources of raw materials and natural resources of that region.

MICHAEL S. KOLEDA, Ph.D. Brown 1970. Centralization vs. decentralization in the supply of public goods.

This thesis analyzes the economic welfare implications of community centralization in the presence of differences in incomes, preferences, and population. Among other conclusions it is found that centralization is always preferable to decentralization when the public good is "pure" and taxation flexible; intercommunity differences in incomes, preferences, and population may alter this conclusion if taxation is inflexible or flexible over a limited range; when centralized supply is inferior to decentralized supply it is always desirable to supply some positive quantity of the public good for the common benefit of both communities.

CHARLES R. LAUDOR, Ph.D. Columbia 1970. The multipurpose River Valley development scheme as a regional growth mechanism.

This study undertakes a comparative economic analysis of actual scheme performance of specified assignments by five multipurpose schemes. In addition to the Tennessee Valley Authority of the United States and the Damodar Valley Authority of India, which serve as examples of schemes in an advanced country and an underdeveloped economy, respectively, three additional schemes are observed: the Sao Francisco Valley Commission of Brazil; the Helmand Valley Authority of Afghanistan; and the Volta River Authority of Ghana.

RAYMOND R. LAUER, Ph.D. Pennsylvania State 1969. An economic evaluation of the work of the Shamokin area industrial corporation.

SHU-JAN LIANG, Ph.D. Oklahoma 1969. A method for projecting county income in Oklahoma.

Examination of Oklahoma county income data for the past two decades reveals that there is a pattern among the counties that persists over a number of years. The stability was found in the percentage of the state total accounted for by each county. On the basis of this information it was found feasible to prepare projections of county income for one or two years. The projections are close to actual estimates for most countries.

STEPHEN E. LILE, Ph.D. Kentucky 1969. Interstate comparisons of family tax burdens.

This study is an attempt to estimate state and local tax burdens in 50 states for one size family (located in each state's largest city) at seven different income levels. The study presents a ranking of states by tax burden significantly different from that based on the conventional taxes per capita and taxes as a percent of personal income measures of burden. This results mainly from the method used to allocate business taxes. These are allocated to residents and non-residents alike depending on the degree of forward shifting assumed. The study also provides an indication of how tax burdens are distributed in different states among different family income levels.

JEROLYN LYLE, Ph.D. Maryland 1970. Racial discrimination in urban labor markets.

This study is a cross-sectional analysis of the relative occupational standing of Negroes in forty-six cities. The regression model indicated that maintaining a low unemployment rate offers occupational advantage to Negro women, but housing desegregation offers more. The relative occupational standing of Negro men in two-digit industries is negatively related to the proportion of industry employees in companies with federal contracts. More effective enforcement of contract compliance by federal authorities is recommended.

HOWARD K. MCCURDY, Ph.D. Cornell 1969. Centralization-decentralization in a context of intergovernmental relations: a study of selected federal projects in Appalachia.

ROBERT G. MCGILLIVRAY, Ph.D. California (Berkeley) 1969. Demand for urban transport.

AHMED R. MANDOUR, Ph.D. Oklahoma 1970. A method of projecting manufacturing activity in Oklahoma, 1970–2020.

This study dealt with the preliminary work needed to project the economy of Oklahoma through 2020 with emphasis on manufacturing activity. It is in line with the national projections undertaken in 1968 by the Office of Business Economics of the U.S. Department of Commerce. The purpose of the study was to determine if sufficient information was available to make reasonable estimates on the state level. The results show that such projections are possible.

DUNCAN R. MILLER, Ph.D. Kentucky 1970. Public investment and regional economic development.

JOHN R. MILLER, Ph.D. California (Los Angeles) 1969. A simulation investigation of the role of due dates in networks of waiting lines.

PAUL T. NELSON, Ph.D. Michigan State 1970. A central-place analysis of selected industrial market structures.

This is a study of the spatial dimension of marketing. The study is a search for order in existing spatial relationships in the marketing of three industrial products through industrial distributors. The research seeks to determine whether Christaller's central-place theory provides a useful framework for the analysis of these relationships. It is concluded that central-place theory does provide a useful basis for analysis which offers insight for managerial decision making and theory development.

NORMAN D. NICHOLSON, Ph.D. Southern California 1969. Cross-sectional analysis of manufacturing production functions in a partitioned "urban/rural" sample space.

This dissertation reports on an analysis of differences in production functions, as they exist, for firms operating in different markets. Essentially, the technique employed involves the development of a dichotomized sample space representing both "urban" and "rural" markets. Analysis of residuals of estimated restricted and unrestricted Cobb-Douglas production functions for 20 manufacturing industries in each market yielded information regarding production efficiency, returns to scale, and underlying market structure for each market. Additional tests were made to determine the existence of agglomeration economies.

LEE D. OLVEY, Ph.D. Harvard 1970. Regional growth and interregional migration: their pattern of interaction.

To explain regional disparities in economic performance, a simultaneous equation model is estimated using employment growth, gross migration, wage and income levels, and other variables for 56 large metropolitan areas for the period 1955–60. Empirical results support the need for gross migration data and for explicit treatment of the simultaneity problem. Wage differentials play a key role in explaining regional employment growth disparities and migration patterns are seen to be highly sensitive to employment growth rates.

THOMAS PETERSON, Ph.D. California (Los Angeles). An economic evaluation of the southern California rapid transit district: its proposed solution to the transportation problem in Los Angeles.

The dissertation analyzes the proposals to solve the transportation problem in the Los Angeles metropolitan area with emphasis on the $2.5 billion, 89 mile rapid transit system which was defeated by the Los Angeles area voters in 1968. Although suggestions are

included as alternative to a rapid transit system, it is concluded that the use of prices to ration the transportation facilities are required if congestion is to be reduced.

ALBERT W. SARGENT, Ph.D. Clark 1970. Estimation of income, employment and tax multipliers for the Worcester standard metropolitan statistical area, by use of an intersectoral flow model.

This study resulted in the construction of three input-output models (rows only) for the Worcester standard metropolitan statistical area. Two transactions tables measure interindustry and interregional trade flows. Direct and indirect requirements per dollar of final demand, with and without the household sector are reported. The model has 29 intermediate and 7 final demand categories. Income, employment, and tax multipliers are estimated for each industry and analyzed.

JAY I. STARK, Ph.D. Michigan 1969. The pattern of resource allocation in education: the Detroit public schools 1940 to 1960.

MITCHELL STENGEL, Ph.D. Harvard 1970. Racial price discrimination in the urban rental housing market.

The existence, sign, and magnitude of racial price discrimination in the urban rental housing market are analyzed through the use of a cross-section regression model of the determinants of gross rent. Comparative studies of the largest SMSA's in the North Central and Southern regions, and of different housing quality submarkets within each region, reveal that the rates of total black population growth and of net black immigration and the degree of residential concentration of blacks within the SMSA are the major determinants of the level of racial price discrimination.

PAUL S. STONE, Ph.D. North Carolina State 1969. Analysis of economic relationships and employment changes in the northeastern region of North Carolina.

A from-to employment flows model was used to estimate county and intercounty employment multipliers for the 35-county region. Changes in employment in each county from 1950–60 were disaggregated into those resulting from changes in export employment and in employment flow coefficients. Average sizes of county employment multipliers were 1.65 in 1950 and 1.77 in 1960. Regional employment multipliers were one-fourth larger in both years.

JACK THORNTON, Ph.D. Missouri 1969. An industry analysis of Audrain County, Missouri.

ROBERT G. TURNER, Ph.D. Kentucky 1969. The locational characteristics of Kentucky and the role of the Kentucky Department of Commerce in the location of new manufacturing plants.

The manufacturing plants in Kentucky were classified into two groups and surveyed by mail questionnaire. One group consisted of plants locating prior to 1964, and the other, plants locating from 1964 to 1967. Material surveyed included the advantages and disadvantages of Kentucky locations, the role of financial assistance and mass advertising programs in inducing plant location, and the impact of the Kentucky Department of Commerce on the location decision.

GARRETT A. VAUGHN, Ph.D. Duke 1970. A comparison of white and nonwhite metropolitan low-income housing.

NORMAN WALZER, Ph.D. Illinois (Urbana) 1970. Economies of scale in municipal police protection.

L. CARL WOEHLCKE, Ph.D. Washington State 1970. Ultra-long-range projections in water resource planning.

ART WRIGHT, Ph.D. Missouri 1970. The determinants of economic growth in the Missouri Ozarks: 1950–65.

Related Disciplines

THOMAS ARMOR, Ph.D. California (Los Angeles) 1969. Peak experiences and sensitivity training.

FRED R. BAHR, D.B.A. George Washington 1970. The expanding role of the Department of Defense as an instrument of social change.

This dissertation represents a detailed study of the expanding role that the defense establishment is playing in the improvement of our society. There are three phases to this study: an evaluation of military participation in internal growth and development; an analysis of current defense programs directed toward alleviating specific social inequities; and an appraisal of proposed domestic action programs to be undertaken by the armed forces.

HARRY D. BAKER, D.B.A. George Washington 1970. The effectiveness of company-sponsored educational programs for nonsupervisory exempt employees.

The purpose of this study was to examine the role and effectiveness of company-sponsored educational programs for nonsupervisory exempt personnel in research- and development-oriented companies. The study primarily involved the evaluation of three hypotheses dealing with: position mobility; job-related needs; and influence of association. Nearly one thousand randomly selected respondents from a division of five R&D-oriented companies in the Baltimore-Washington metropolitan area completed anonymous questionnaires. A modified chi-square method of statistical analysis was used in testing the hypotheses.

ROBERT J. BOLDIN, Ph.D. Pennsylvania 1970. A man-machine planning model for a university admissions function.

GEORGE DIEHR, Ph.D. California (Los Angeles) 1969. An investigation of computational algorithms for aggregation problems.

ROBERT H. DOKTOR, Ph.D. Stanford 1970. The development and mapping of certain cognitive styles of problem solving.

AHMED EL-JACK, Ph.D. California (Los Angeles) 1969. Management development programs in the Sudan: a comparative study.

PAUL C. ERGLER, D.B.A. George Washington 1969. A study of contract engineering as used in technical manpower management in the Baltimore, Maryland-Washington, D.C. area.

Contract engineering is the renting of personnel to perform technical services. The research objective was to assess the concept's value. Subsidiary questions involved: supply and demand, reasons for use, quality of contract versus permanent personnel, operating problems, costs, future outlook, disadvantages, and effect of client size. The concept was concluded to be valuable as a staffing tool if use were based upon weighing alternatives and situational factors by discriminating managers.

SALEH FARID, Ph.D. California (Los Angeles) 1969. The organizational role of the board of directors of state enterprises in Egypt.

C. W. GETZ, D.B.A. George Washington 1970. An analysis of management of the data resource.

Data is a basic resource of the firm in the same meaning as the other physical, financial, and human resources. Thus, the functions of management—planning, organizing, coordinating, commanding, and control—can and should be applied to management of the data resource. It is the computer that has focused current attention on data. Data is expensive. To remain competitive, firms must devote substantial resources and attention to management of their data resources.

NEIL J. HUMPHREYS, Ph.D. Pennsylvania 1970. Understanding the office of the elementary principal in the school district of Philadelphia.

This dissertation presents an organizational study of the elementary principalship in a large city school system. Role perceptions of the elementary principal population and role expectations of three counter position groups are compared with actual role performance of four principals. Conclusions indicate that principal observed behavior is primarily a function of the incumbents interpretation of the office. The three counter position groups do not hold well-defined expectations for principal role behavior.

THOMAS KLINCK, Ph.D. California (Los Angeles) 1969. Assessing organizational effectiveness and developing a strategy for change for a community action agency board of directors.

ROSS E. LANSER, Ph.D. Stanford 1969. Visible traits of boards of directors of new enterprise.

JAMES P. MCNAUL, Ph.D. Stanford 1970. Behavioral patterns among professionals in a research and development environment.

ROBERT L. MANDEVILLE, D.B.A. George Washington 1969. A comparative study of self-perceived decision process influence exerted by NASA program managers, NASA project managers, and industry managers on NASA-funded projects.

A study of activities performed by National Aeronautics and Space Administration (NASA) program and project managers, and their counterparts in private industry who manage NASA-funded research and development as perceived by the managers themselves. The academic significance of this research lies in the establishment of a base set of data comparing and contrasting the most time-consuming and important R&D project activities as perceived by the three groups of managers.

ALLEN R. MICKELSON, Ph.D. Virginia 1969. Title I of the higher education act: a case study of bureaucracy.

LAWRENCE T. PINFIELD, Ph.D. Stanford 1970. Interaction patterns of members of international offshoot organizations.

PRESTON PROBASCO, Ph.D. Wisconsin 1969. Social feedback factors in vocal shadowing.

PETER RAYNOLDS, Ph.D. California (Los Angeles) 1969. The projective-differential: a general purpose inkblot technique for studying denotable objects.

MILTON SHAW, Ph.D. Pennsylvania 1969. The ability to shape attitudes: an indicator of the effectiveness of leadership.

It is demonstrated that the ability to shape or transform the attitudes of others is a significant factor in determining leadership effectiveness. In the theory evolved, leadership, defined as the possession of the ability to shape or transform the attitudes of subordinates, is the active or independent variable. It is directly related to the intervening variable: attitudes of subordinates. These attitudes are in turn functionally related to the group's effectiveness in achieving a defined goal.

DENNIS P. SLEVIN, Ph.D. Stanford 1969. A mathematical model of "standing pat" behavior.

JEREMIAH M. SULLIVAN, Ph.D. Princeton 1970. Estimation of childhood mortality conditions from childhood survival statistics.

This study develops two models each of which is designed to estimate the probability of surviving from birth to ages 2, 3 and 5. The models require as input, data collected at a single point in time via survey or census methods, on the proportion surviving of the children ever born to respondent women. The models differ in being designed to employ data classified by

age intervals or, alternatively, marriage duration intervals of female respondents. Empirical testing of the models was also undertaken in the study. Estimates from each model based on data collected by the Spanish decennial censuses of 1930 and 1940 are evaluated and interpreted with the aid of mortality statistics from the Spanish Registration System.

SANFORD TEMKIN, Ph.D. Pennsylvania 1969. A cost-effectiveness approach to improving resource allocations for school systems.

This dissertation focuses on the economic situation of school systems where demand for resources far exceeds supply. The study's purposes were: 1) develop a comprehensive theory of cost-effectiveness decision-making; 2) operationalize relevant portions of the theory to evaluate ongoing operations of a school system; and 3) based on what presently exists develop alternative ways to change. Sensitivity analysis is used to examine output responses as a function of parameter modifications.

ROBERT TURRILL, Ph.D. California (Los Angeles) 1969. Use of the semantic differential in relating perceived self-organizational environment similarity to the meaning of membership in a bureaucracy.

NATIONAL REGISTRY FOR ECONOMISTS

The National Registry for Economists was established in January, 1966, to provide a centralized nationwide clearinghouse for economists on a year-round basis. It is located in the Chicago Professional Placement Office of the Illinois State Employment Service and is staffed by experienced placement personnel, operating under the guidance and direction of Regional and National Bureau of Employment Security Professional Placement officials, and in cooperation with the American Economic Association. It is a free service. There are no registration, referral, or placement fees. Application and order forms used in the Registry are available upon request from the: National Registry for Economists, Professional Placement Center, 208 South La Salle Street, Chicago, Illinois 60604.

AMERICAN ECONOMIC ASSOCIATION

VACANCIES AND APPLICATIONS

The Association is glad to render service to applicants who wish to make known their availability for positions in the field of economics and to administrative officers of colleges and universities and others who are seeking to fill vacancies in the field of economics.

The officers of the Association take no responsibility for making a selection among the applicants or following up the results. The Secretary's office will merely afford a central point for forwarding inquiries, and the *Review* will publish in this section a brief description of vacancies announced and of applications submitted (with necessary editorial changes). Since the Association has no other way of knowing whether or not this section is performing a real service, the Secretary would appreciate receiving notification of appointments made as a result of these announcements. Those submitting such announcements have the option of publishing either name and address or a key number in the listing. Inquiries about a listing with a key number should refer to it specifically and be mailed to the Secretary's office. Resumes and application blanks are not supplied by the American Economic Association. The Association will only forward inquiries and resumes to the proper party for their consideration. Deadlines for the four issues of the *Review* are January 1, April 1, July 1, and October 1.

Communications should be addressed to: The Secretary, American Economic Association, 809 Oxford House, 1313 21st Avenue South, Nashville, Tennessee 37212.

Vacancies

Fishery economists: Wide variety of economics research, ranging from international agreements, quotas, and tariffs to price analyses, business management of firms, cost-benefit analyses, and whole field of the economics of natural resources. Positions are in the federal Civil Service at Grades GS-9 ($9,320) to GS-14 ($18,531). Basic requirements are Ph.D. or master's in economics or agricultural economics; training in international, natural resource, and/or quantitative economics would be helpful. Positions are located at the University of Maryland, in Washington, D.C., and field positions. Civil Service Commission Form 171 (Application for Federal Employment) should be sent to: Personnel Office, Bureau of Commercial Fisheries, U.S. Department of the Interior, 18th and C Street, N.W., Washington, D.C., 20240.

Economist: Opening in the Department of State Planning. Master's degree in economics and six years of experience required. Additional graduate study may be substituted for two years of the required experience. Background in quantitative research methods desirable. Starting salary $13,739 effective September 1, 1970, with a maximum of $18,049 reached in six years. Write: Vladimir Wahbe, Secretary of State Planning, Department of State Planning, 301 West Preston Street, Baltimore, Maryland, 21201.

Business administration: Applications sought at all rank levels for rapidly expanding program, favorably located in Vancouver metropolitan area. Both undergraduate and graduate work, but emphasis on latter. Integrated with strong economics program; well established at B.A.-Ph.D. levels. Additional specialists sought in behavioral aspects of management, organization theory, finance, managerial economics, marketing, information systems, and business policy. Doctorate required. Write to: Dr. John P. Herzog, Acting Chairman, Department of Economics and Commerce, Simon Fraser University, Burnaby 2, British Columbia, Canada.

Senior appointment in economics: Newer university, attractively located in Vancouver metropolitan area, seeks candidates with notable academic record to supervise doctoral research programs and teach at various levels in substantial economics program with well-qualified students. Write to: Dr. John P. Herzog, Acting Chairman, Department of Economics and Commerce, Simon Fraser University, Burnaby 2, British Columbia, Canada.

Staff economist—systems analyst: The Center for Naval Analyses, operated under contract with the University of Rochester, has three-four openings for Ph.D. economists. CNA is engaged in a broad spectrum of research involving operations research and systems analysis for the U.S. Navy and government agencies. CNA analyses provide bases for decisions in the planning and budgeting of military forces and government operations. This involves the application of economic principles to resource allocation problems relating to the sizing, mix, procurement, and use of alternative systems. Responsibilities include defining and structuring problems of choice, identifying alternatives, analysis of alternatives and constraints, illuminating relative costs and benefits. We have recently studied: the naval ship procurement process, cost and effectiveness of land-based and sea-based tactical aviation, trade-offs between aircraft procurement and maintenance, and the supply of labor to the military in the absence of a draft. We now plan to study: the organization and operation of the Navy resource allocation process, the allocation of air space between Navy and other users, the design of strategic forces, the costs and benefits of military medical systems, and a worldwide review of the naval aviation shore establishment. Time is available for professional development and research directed toward improving and/or developing analytical techniques and decision processes. Starting salaries range from $16,000 to $24,000, depending on education and experience. Staff members are eligible for merit increases at least annually. Write to: Mr. Terry E. Harris, Professional Staffing, Center for Naval Analyses, 1401 Wilson Boulevard, Arlington, Virginia, 22209. CNA is an Equal Opportunity Employer.

Industrial relations: A Ph.D. in industrial relations or economics to join the faculty in the expanding graduate (M.A.) program in industrial relations at a private college in the East, starting in January, 1971. Rank and salary open. P365

Economics: The United Nations has a continuing need for economists for challenging research and operational posts at Headquarters, New York, in Geneva, Vienna, and in developing countrie throughout the world. These posts cover the ful field of economic specialization (econometrics, fiscal policy, foreign trade, finance, agriculture, mining, water resources, industrial management, development, urban planning, etc.), at all levels of responsibility and of varying duration, permanent or short term. Representative of the posts now under recruitment are the following. The gross salary for these positions ranges between $12,380 and $20,490 ($9,666-$15,118 tax free), exclusive of dependence and post allowances which may amount to an additional several thousand dollars. (1) ECE, Geneva: International economists with several years of research experience and particular qualifications in East-West European trade; knowledge of French or Russian required. (2) ECLA, Santiago: Industrial economist with technical-economic research experience and practical experience in Latin American industrial sectors; e.g., textiles and metal transforming; Spanish required. U.S. citizens interested in obtaining additional information about these as well as other UN employment possibilities are requested to send a résumé to: Office of International Organization Recruitment, Room 4336A (IO/IR), U.S. Department of State, Washington, D.C., 20520

Regional or urban economics, operations research: September or December, 1970; Ph.D. required. The Management Sciences-Operations Research Department in the Technological Institute of a Big Ten school has a vacancy. The candidate should have a Ph.D. in economics (theory, urban and regional) with training in operations research or a Ph.D. in operations research with extensive graduate work in economics. Teaching would include undergraduate courses in managerial economics and graduate courses in the applications of operations research and economic analysis to urban problems, plant investment, and location problems, and resource allocation problems. As a member of the newly formed Urban Systems Engineering Center, the candidate would have the opportunity and freedom to direct his research and his students toward relevant problems. Appointments can be made at all levels, depending upon the candidate's qualifications and experience. Teaching load will be three to five quarter courses for the academic year. P367

Economist: Ph.D., economics, with research and publication experience. Position to encompass quantitative research in the Center for Business and Economic Research and teaching at the grad-

uate or undergraduate level. Rank of associate or full professor with salary dependent upon qualifications. Please send résumé to: Professor J. F. Niss, Chairman, Department of Economics, Western Illinois University, Macomb, Illinois, 61455.

Economist, econometrician, air transport research director: Air transport industry association is establishing new position, Director of Economic Systems Research, to develop the quantitative and/or scientific approach to the many economic problems faced by scheduled air carriers. Advanced degree in economics and knowledge of econometrics are required. Location, Washington, D.C. Write to: Golightly & Company International, Inc., 1 Rockefeller Plaza, New York, N.Y., 10020.

Economics: Teachers for Department of Economics, 1971-72. Candidates with experience in academic teaching and research may apply to Academic Secretary. Detailed vita, list of publications, three references required. Knowledge of the Hebrew language desirable. University will assist accepted candidates with housing and transportation. Applications accepted until December 31, 1970. Write to: University of Haifa, Mount Carmel, Haifa, Israel.

Economist: The National Association of Supervisors of State Banks (NASSB) has a vacancy in the position of research associate to be responsible for continuing research program under direction of Executive Vice-President. NASSB is engaged in a broad range of bank and bank supervisory research including bank structure and performance, bank code revision, examination procedures, and background data for congressional testimony and various litigation and maintenance of comparative statistical indices. Broad opportunity to initiate research projects. Prefers candidates who have recently completed Ph.D. or D.B.A. but will consider doctoral candidates (M.S.-M.B.A. minimum). Salary competitive. Write to: Dr. Harry P. Guenther, Executive Vice President-Economist, NASSB, 1101 17th Street, N.W., Washington, D.C., 20036.

Marketing, accounting and finance, engineering economy: Positions for associate professor of marketing and of accounting and finance and assistant professor of engineering economy. Ph.D. plus university and/or industrial experience. Teaching and research in areas of specialization. Salary competitive. Send résumé to: Chairman, Department of Engineering Management, University of Missouri, Rolla, Missouri, 65401.

Business administration: Rapidly growing branch of University of Alaska, offering B.B.A., B.A. in Economics, and M.B.A., requires D.B.A. or Ph.D. in business administration or dissertation stage. Nine-hour teaching load covering graduate and undergraduate programs plus supervision of graduate students. Rapidly expanding program in Anchorage, the largest population center in state. Research opportunities, summer teaching, plus chance to help formulate departmental programs and policies. Write to: Dr. R. Maurice Jones, M.B.A. Program, University of Alaska, Southcentral Regional Center, 1820 West Northern Lights Boulevard, Anchorage, Alaska, 99503.

Accounting: Two accounting positions open for September, 1971. Rank can be from instructor to professor. Salaries are competitive and will be commensurate with education and experience. Interested parties may write or phone at our expense. Contact: Dr. Lowell Chapman, Head, Department of Accounting, Ferris State College, Big Rapids, Michigan, 49307. Telephone: 616-796-8359, Ext. 526.

Economist: Opening for September, 1971, in small teaching-oriented liberal arts college. Prefers Ph.D. and some teaching experience. Fields needed include American economic history, micro, labor, money and banking. Competitive salary depending on qualifications. Generous fringe benefits and consulting opportunities in local "Model Cities" program. Please send résumé to: Dr. Dennis J. Mahar, Acting Chairman, Pikeville College, Pikeville, Kentucky, 41501.

Economics: Ph.D. or D.B.A. preferred to fill positions in recently established Department of Economics offering a major in economics to teach principles of economics and undergraduate "mathematical" economics and/or first-year accounting. Rank and salary: assistant professor to professor, $9,000 to $16,000, depending on degrees and experience. Write to: Department of Economics, Edinboro State College, Edinboro, Pennsylvania, 16412.

Department chairman: A fifteen-man Department of Economics with undergraduate and graduate programs requires a man with Ph.D. and teaching and research experience as department chairman. Salary commensurate with academic qualifications and experience. Submit résumé to: J. W. Skinner, Dean, College of Social Science, University of Guelph, Guelph, Ontario, Canada.

Economics: Several young economists with doctorate completed for September, 1971. Various specialties but a strong theoretical or quantitative orientation expected. Knowledge of computer utilization desired of one new faculty member. Good teaching and some publication required for tenure and promotion. Favorable location in interior of northern California. Write to: John A. Orr, Chairman, Department of Economics, Chico State College, Chico, California, 95926.

Economists: Applications are invited for two or three appointments in January or September, 1971. Ph.D. required for professorial ranks. Please send résumé to: Dr. I. Ghurani, Chairman,

Department of Economics, Pahlavi University, Shiraz, Iran.

Business administration: Applications are invited for one or two appointments in accounting and in general business. Ph.D. required for professorial ranks. Please send résumé to: Chairman, Department of Economics, Pahlavi University, Shiraz, Iran.

Academic dean: The Austin Dunham Barney School of Business and Public Administration, University of Hartford, is seeking a dean with innovative approaches to both graduate and undergraduate education in the fields of management and economics. Experience in business or government and university teaching experience preferred. Ph.D. required. Starting salary between $20-$25,000. Beginning January 1, 1971. Please contact: Professor Edward F. McDonough, Austin Dunham Barney School of Business and Public Administration, University of Hartford, 200 Bloomfield Avenue, West Hartford, Connecticut, 06117.

Finance and management: National Development Corporation, responsible for promoting development prospects in Tanzania, has vacancy for Group Loan Officer in Office of Planning and Finance. The Group Loan Officer ensures that the Corporation's loans are made in accordance with established policies and standards of financial prudence; manages the deployment of short-term funds between Group companies; and is responsible for providing continual guidance to NDC top management on prevailing conditions in worldwide money markets. Salary negotiable depending on qualifications. Write to: George Kahama, Managing Director, National Development Corporation, Development House, Azania Front, P.O. Box 2669, Dar es Salaam, Tanzania.

Economics, management science: Two new positions for 1971-72 with primary specialties in the areas of behavioral economics, finance, simulation models for decision making and economic theory. Appointments will be made at the assistant and associate level. Ph.D. or near Ph.D. required. Minimum salary A.A.U.P.-A. Address applications to: Dr. David A. Martin, Chairman, Department of Economics, State University of New York, College at Geneseo, New York, 14454.

Marketing: Rapidly growing private, nonsectarian, coeducational college seeks chairman for expanding Marketing Department. Excellent potential for growth in pleasant work and residential environment on new suburban campus near Princeton, New Jersey. Applicant should have evidence of scholarly research and publication record. Training in behavioral sciences or quantitative methods and computer experience are desired. The applicant must have Ph.D. or D.B.A. with administrative and teaching experience. Teaching load includes undergraduate and graduate courses in areas of personal interest and training. Rank and salary competitive. Inquiries should be sent to: D. N. Dertouzos, Dean, School of Business Administration, Rider College, Trenton, New Jersey, 08602.

Business administration: The Business Administration Department of Mankato State College is seeking applicants for the position of department chairman. Applicant must possess an earned doctorate, teaching experience, and administrative supervisory or business experience is desirable. Salary $20,000-$22,500. Send application to: Dr. Morgan I. Thomas, Dean, School of Business, Mankato State College, Mankato, Minnesota, 56001.

Business administration: Michigan church-related (Methodist) four-year liberal arts college seeks M.A. plus, M.B.A. plus, Ph.D., or D.B.A. for teaching general business and marketing courses in 1970. Flexible course assignments. Rank depending on applicant's qualifications. Beginning salary range: $8,000-$13,000. Generous fringe benefits. P368

Economist: Attractive openings in Washington, D.C., for individuals with Ph.D. or exceptional graduate school record without completed dissertation for position (two-three days per week) doing research on social welfare problems. Background in public finance and microeconomics preferred. Ideal for woman who wishes to combine domesticity and important work or faculty member who wishes to combine teaching with policy analysis. P369

Economics—general competence, interest in both undergraduate and graduate teaching emphasizing development economics. One or more specialties: theory, history, systems, labor, finance, money and banking, international trade, accounting, econometrics policy, computer application.
Minimum four years post-doctoral experience English-medium institution, excellent teaching recommendations, professional and personal adaptability. Compensation comparable average U.S. institutions or willing sub-contract, two year appointments customary. Write Dean of Faculties, American University in Cairo, at NY Office, 866 UN Plaza, NY, NY 10017.

Economic Planning and Development—for undergraduate and graduate programs. Extensive and varied teaching and practical experience desired. General competence and additional specialties welcome.
Minimum four years post-doctoral experience English-medium institution, excellent teaching recommendations, professional and personal adaptability. Compensation comparable average U.S. institutions or willing sub-contract, two year appointments customary. Write Dean of Faculties, American University in Cairo, at NY Office, 866 UN Plaza, NY, NY 10017.